FOR STUDENTS

Solutions to Odd-Numbered End-of-Chapter Exercises

Chapter 2
Review of Probability

■ Solutions to Exercises

1. (a) Probability distribution function for Y

Outcome (number of heads)	$Y = 0$	$Y = 1$	$Y = 2$
probability	0.25	0.50	0.25

(b) Cumulative probability distribution function for Y

Outcome (number of heads)	$Y < 0$	$0 \le Y < 1$	$1 \le Y < 2$	$Y \ge 2$
Probability	0	0.25	0.75	1.0

(c) $\mu_Y = E(Y) = (0 \times 0.25) + (1 \times 0.50) + (2 \times 0.25) = 1.00$

Using Key Concept 2.3: $\text{var}(Y) = E(Y^2) - [E(Y)]^2$, and

$E(Y^2) = (0^2 \times 0.25) + (1^2 \times 0.50) + (2^2 \times 0.25) = 1.50$

so that $\text{var}(Y) = E(Y^2) - [E(Y)]^2 = 1.50 - (1.00)^2 = 0.50.$

3. For the two new random variables $W = 3 + 6X$ and $V = 20 - 7Y$, we have:

(a)

$$E(V) = E(20 - 7Y) = 20 - 7E(Y) = 20 - 7 \times 0.78 = 14.54,$$
$$E(W) = E(3 + 6X) = 3 + 6E(X) = 3 + 6 \times 0.70 = 7.2.$$

(b)

$$\sigma_W^2 = \text{var}(3 + 6X) = 6^2 \cdot \sigma_X^2 = 36 \times 0.21 = 7.56,$$
$$\sigma_V^2 = \text{var}(20 - 7Y) = (-7)^2 \cdot \sigma_Y^2 = 49 \times 0.1716 = 8.4084.$$

(c)

$$\sigma_{WV} = cov(3 + 6X, 20 - 7Y) = 6(-7)cov(X, Y) = -42 \times 0.084 = -3.528$$
$$cor(W, V) = \frac{\sigma_{WV}}{\sigma_W \sigma_V} = \frac{-3.528}{\sqrt{7.56 \times 8.4084}} = -0.4425.$$

5. Let X denote temperature in °F and Y denote temperature in °C. Recall that $Y = 0$ when $X = 32$ and $Y = 100$ when $X = 212$; this implies $Y = (100/180) \times (X - 32)$ or $Y = -17.78 + (5/9) \times X.$ Using Key Concept 2.3, $\mu_X = 70°\text{F}$ implies that $\mu_Y = -17.78 + (5/9) \times 70 = 21.11°\text{C},$ and $\sigma_X = 7°\text{F}$ implies $\sigma_Y = (5/9) \times 7 = 3.89°\text{C}.$

7. Using obvious notation, $C = M + F;$ thus $\mu_C = \mu_M + \mu_F$ and $\sigma_C^2 = \sigma_M^2 + \sigma_F^2 + 2\,\text{cov}(M, F).$ This implies

(a) $\mu_C = 40 + 45 = \$85{,}000$ per year.

(b) $cor(M, F) = \frac{Cov(M, F)}{\sigma_M \sigma_F},$ so that $Cov(M, F) = \sigma_M \sigma_F cor(M, F).$ Thus $Cov(M, F) = 12 \times 18 \times 0.80 = 172.80,$ where the units are squared thousands of dollars per year.

(c) $\sigma_C^2 = \sigma_M^2 + \sigma_F^2 + 2\,\text{cov}(M, F),$ so that $\sigma_C^2 = 12^2 + 18^2 + 2 \times 172.80 = 813.60,$ and $\sigma_C = \sqrt{813.60} = 28.524$ thousand dollars per year.

(d) First you need to look up the current Euro/dollar exchange rate in the Wall Street Journal, the Federal Reserve web page, or other financial data outlet. Suppose that this exchange rate is e (say $e = 0.80$ euros per dollar); each 1\$ is therefore with eE. The mean is therefore $e\mu_C$ (in units of thousands of euros per year), and the standard deviation is $e\sigma_C$ (in units of thousands of euros per year). The correlation is unit-free, and is unchanged.

9.

		Value of *Y*					**Probability Distribution of *X***
		14	**22**	**30**	**40**	**65**	
Value of *X*	**1**	0.02	0.05	0.10	0.03	0.01	0.21
	5	0.17	0.15	0.05	0.02	0.01	0.40
	8	0.02	0.03	0.15	0.10	0.09	0.39
Probability distribution of *Y*		0.21	0.23	0.30	0.15	0.11	1.00

(a) The probability distribution is given in the table above.

$$E(Y) = 14 \times 0.21 + 22 \times 0.23 + 30 \times 0.30 + 40 \times 0.15 + 65 \times 0.11 = 30.15$$
$$E(Y^2) = 14^2 \times 0.21 + 22^2 \times 0.23 + 30^2 \times 0.30 + 40^2 \times 0.15 + 65^2 \times 0.11 = 1127.23$$
$$\text{Var(Y)} = E(Y^2) - [E(Y)]^2 = 218.21$$
$$\sigma_Y = 14.77$$

(b) Conditional Probability of $Y|X = 8$ is given in the table below

Value of *Y*				
14	**22**	**30**	**40**	**65**
0.02/0.39	0.03/0.39	0.15/0.39	0.10/0.39	0.09/0.39

$$E(Y|X=8) = 14 \times (0.02/0.39) + 22 \times (0.03/0.39) + 30 \times (0.15/0.39)$$
$$+ 40 \times (0.10/0.39) + 65 \times (0.09/0.39) = 39.21$$
$$E(Y^2|X=8) = 14^2 \times (0.02/0.39) + 22^2 \times (0.03/0.39) + 30^2 \times (0.15/0.39)$$
$$+ 40^2 \times (0.10/0.39) + 65^2 \times (0.09/0.39) = 1778.7$$
$$\text{Var}(Y) = 1778.7 - 39.21^2 = 241.65$$
$$\sigma_{Y|X=8} = 15.54$$

(c) $E(XY) = (1 \times 14 \times 0.02) + (1 \times 22 : 0.05) + \cdots (8 \times 65 \times 0.09) = 171.7$

$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 171.7 - 5.33 \times 30.15 = 11.0$

$\text{Corr}(X, Y) = \text{Cov}(X, Y)/(\sigma_X \sigma_Y) = 11.0/(5.46 \times 14.77) = 0.136$

11. (a) 0.90
    (b) 0.05
    (c) 0.05
    (d) When $Y \sim \chi^2_{10}$, then $Y/10 \sim F_{10,\infty}$.
    (e) $Y = Z^2$, where $Z \sim N(0,1)$, thus $\Pr(Y \le 1) = \Pr(-1 \le Z \le 1) = 0.32$.

13. (a) $E(Y^2) = \text{Var}(Y) + \mu_Y^2 = 1 + 0 = 1; E(W^2) = \text{Var}(W) + \mu_W^2 = 100 + 0 = 100.$
    (b) *Y* and *W* are symmetric around 0, thus skewness is equal to 0; because their mean is zero, this means that the third moment is zero.
    (c) The kurtosis of the normal is 3, so $3 = \frac{E(Y-\mu_Y)^4}{\sigma_Y^\$}$; solving yields $\text{E}(Y^4) = 3$; a similar calculation yields the results for *W*.
    (d) First, condition on $X = 0$, so that $S = W$:

$$E(S|X=0) = 0; E(S^2|X=0) = 100, E(S^3|X=0) = 0, E(S^4|X=0) = 3 \times 100^2.$$

Similarly,

$$E(S|X=1) = 0; E(S^2|X=1) = 1, E(S^3|X=1) = 0, E(S^4|X=1) = 3.$$

From the large of iterated expectations

$$E(S) = E(S|X=0) \times \Pr(X=0) + E(S|X=1) \times \Pr(X=1) = 0$$
$$E(S^2) = E(S^2|X=0) \times \Pr(X=0) + E(S^2|X=1) \times \Pr(X=1) = 100 \times 0.01 + 1 \times 0.99 = 1.99$$
$$E(S^3) = E(S^3|X=0) \times \Pr(X=0) + E(S^3|X=1) \times \Pr(X=1) = 0$$
$$E(S^4) = E(S^4|X=0) \times \Pr(X=0) + E(S^4|X=1) \times \Pr(X=1) = 3 \times 100^2 \times 0.01 + 3 \times 1 \times 0.99 = 302.97$$

(e) $\mu_S = E(S) = 0$, thus $E(S - \mu_S)^3 = E(S^3) = 0$ from part d. Thus skewness = 0.

Similarly, $\sigma_S^2 = E(S - \mu_S)^2 = E(S^2) = 1.99$, and $E(S - \mu_S)^4 = E(S^4) = 302.97$.

Thus, $\text{kurtosis} = 302.97/(1.99^2) = 76.5$

15. (a)

$$\begin{aligned}\Pr(9.6 \le \bar{Y} \le 10.4) &= \Pr\left(\frac{9.6-10}{\sqrt{4/n}} \le \frac{\bar{Y}-10}{\sqrt{4/n}} \le \frac{10.4-10}{\sqrt{4/n}}\right)\\ &= \Pr\left(\frac{9.6-10}{\sqrt{4/n}} \le Z \le \frac{10.4-10}{\sqrt{4/n}}\right)\end{aligned}$$

where $Z \sim N(0, 1)$. Thus,

(i) $n = 20;\ \Pr\left(\frac{9.6-10}{\sqrt{4/n}} \le Z \le \frac{10.4-10}{\sqrt{4/n}}\right) = \Pr(-0.89 \le Z \le 0.89) = 0.63$

(ii) $n = 100;\ \Pr\left(\frac{9.6-10}{\sqrt{4/n}} \le Z \le \frac{10.4-10}{\sqrt{4/n}}\right) = \Pr(-2.00 \le Z \le 2.00) = 0.954$

(iii) $n = 1000;\ \Pr\left(\frac{9.6-10}{\sqrt{4/n}} \le Z \le \frac{10.4-10}{\sqrt{4/n}}\right) = \Pr(-6.32 \le Z \le 6.32) = 1.000$

(b)

$$\begin{aligned}\Pr(10-c \le \bar{Y} \le 10+c) &= \Pr\left(\frac{-c}{\sqrt{4/n}} \le \frac{\bar{Y}-10}{\sqrt{4/n}} \le \frac{c}{\sqrt{4/n}}\right)\\ &= \Pr\left(\frac{-c}{\sqrt{4/n}} \le Z \le \frac{c}{\sqrt{4/n}}\right).\end{aligned}$$

As n get large $\frac{c}{\sqrt{4/n}}$ gets large, and the probability converges to 1.

(c) This follows from (b) and the definition of convergence in probability given in Key Concept 2.6.

17. $\mu_Y = 0.4$ and $\sigma_Y^2 = 0.4 \times 0.6 = 0.24$

(a) (i) $P(\bar{Y} \ge 0.43) = \Pr\left(\frac{\bar{Y}-0.4}{\sqrt{0.24/n}} \ge \frac{0.43-0.4}{\sqrt{0.24/n}}\right) = \Pr\left(\frac{\bar{Y}-0.4}{\sqrt{0.24/n}} \ge 0.6124\right) = 0.27$

(ii) $P(\bar{Y} \le 0.37) = \Pr\left(\frac{\bar{Y}-0.4}{\sqrt{0.24/n}} \le \frac{0.37-0.4}{\sqrt{0.24/n}}\right) = \Pr\left(\frac{\bar{Y}-0.4}{\sqrt{0.24/n}} \le -1.22\right) = 0.11$

(b) We know $\Pr(-1.96 \le Z \le 1.96) = 0.95$, thus we want n to satisfy $0.41 = \frac{0.41-0.4}{\sqrt{0.24/n}} > -1.96$ and $\frac{0.39-0.4}{\sqrt{0.24/n}} < -1.96$. Solving these inequalities yields $n \ge 9220$.

19. (a)

$$\begin{aligned}\Pr(Y = y_j) &= \sum_{i=1}^{l} \Pr(X = x_i, Y = y_j)\\ &= \sum_{i=1}^{l} \Pr(Y=y_j|X=x_i)\Pr(X=x_i)\end{aligned}$$

(b)

$$E(Y) = \sum_{j=1}^{k} y_j \Pr(Y = y_j) = \sum_{j=1}^{k} y_j \sum_{i=1}^{l} \Pr(Y = y_j | X = x_i) \Pr(X = x_i)$$
$$= \sum_{i=1}^{l} \left(\sum_{j=1}^{k} y_j \Pr(Y = y_j | X = x_i) \right) \Pr(X = x_i)$$
$$= \sum_{i=1}^{l} E(Y | X = x_i) \Pr(X = x_i).$$

(c) When X and Y are independent,

$$\Pr(X = x_i, Y = y_j) = \Pr(X = x_i)\Pr(Y = y_j),$$

so

$$\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)]$$
$$= \sum_{i=1}^{l} \sum_{j=1}^{k} (x_i - \mu_X)(y_j - \mu_Y) \Pr(X = x_i, Y = y_j)$$
$$= \sum_{i=1}^{l} \sum_{j=1}^{k} (x_i - \mu_X)(y_j - \mu_Y) \Pr(X = x_i) \Pr(Y = y_j)$$
$$= \left(\sum_{i=1}^{l} (x_i - \mu_X) \Pr(X = x_i) \right) \left(\sum_{j=1}^{k} (y_j - \mu_Y) \Pr(Y = y_j \right)$$
$$= E(X - \mu_X) E(Y - \mu_Y) = 0 \times 0 = 0,$$
$$cor(X, Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{0}{\sigma_X \sigma_Y} = 0.$$

21. (a)

$$E(X - \mu)^3 = E[(X - \mu)^2 (X - \mu)] = E[X^3 - 2X^2\mu + X\mu^2 - X^2\mu + 2X\mu^2 - \mu^3]$$
$$= E(X^3) - 3E(X^2)\mu + 3E(X)\mu^2 - \mu^3 = E(X^3) - 3E(X^2)E(X) + 3E(X)[E(X)]^2 - [E(X)]^3$$
$$= E(X^3) - 3E(X^2)E(X) + 2E(X)^3$$

(b)

$$E(X - \mu)^4 = E[(X^3 - 3X^2\mu + 3X\mu^2 - \mu^3)(X - \mu)]$$
$$= E[X^4 - 3X^3\mu + 3X^2\mu^2 - X\mu^3 - X^3\mu + 3X^2\mu^2 - 3X\mu^3 + \mu^4]$$
$$= E(X^4) - 4E(X^3)E(X) + 6E(X^2)E(X)^2 - 4E(X)E(X)^3 + E(X)^4$$
$$= E(X^4) - 4[E(X)][E(X^3)] + 6[E(X)]^2[E(X^2)] - 3[E(X)]^4$$

23. X and Z are two independently distributed standard normal random variables, so

$$\mu_X = \mu_Z = 0, \sigma_X^2 = \sigma_Z^2 = 1, \sigma_{XZ} = 0.$$

(a) Because of the independence between X and Z, $\Pr(Z = z | X = x) = \Pr(Z = z)$, and $E(Z|X) = E(Z) = 0$. Thus $E(Y|X) = E(X^2 + Z|X) = E(X^2|X) + E(Z|X) = X^2 + 0 = X^2$.

(b) $E(X^2) = \sigma_X^2 + \mu_X^2 = 1,$ and $\mu_Y = E(X^2 + Z) = E(X^2) + \mu_Z = 1 + 0 = 1.$

(c) $E(XY) = E(X^3 + ZX) = E(X^3) + E(ZX).$ Using the fact that the odd moments of a standard normal random variable are all zero, we have $E(X^3) = 0.$ Using the independence between X and $Z,$ we have $E(ZX) = \mu_Z \mu_X = 0.$ Thus $E(XY) = E(X^3) + E(ZX) = 0.$

(d)

$$\begin{aligned} Cov(XY) &= E[(X - \mu_X)(Y - \mu_Y)] = E[(X - 0)(Y - 1)] \\ &= E(XY - X) = E(XY) - E(X) \\ &= 0 - 0 = 0. \end{aligned}$$

$$cor(X, Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{0}{\sigma_X \sigma_Y} = 0.$$

Chapter 3
Review of Statistics

■ Solutions to Exercises

1. The central limit theorem suggests that when the sample size (n) is large, the distribution of the sample average ($\bar{Y}$) is approximately $N\left(\mu_Y, \sigma_{\bar{Y}}^2\right)$ with $\sigma_{\bar{Y}}^2 = \frac{\sigma_Y^2}{n}$. Given a population $\mu_Y = 100$, $\sigma_Y^2 = 43.0$, we have

   (a) $n = 100$, $\sigma_{\bar{Y}}^2 = \frac{\sigma_Y^2}{n} = \frac{43}{100} = 0.43$, and

$$\Pr(\bar{Y} < 101) = \Pr\left(\frac{\bar{Y} - 100}{\sqrt{0.43}} < \frac{101 - 100}{\sqrt{0.43}}\right) \approx \Phi(1.525) = 0.9364.$$

   (b) $n = 64$, $\sigma_{\bar{Y}}^2 = \frac{\sigma_Y^2}{64} = \frac{43}{64} = 0.6719$, and

$$\Pr(101 < \bar{Y} < 103) = \Pr\left(\frac{101 - 100}{\sqrt{0.6719}} < \frac{\bar{Y} - 100}{\sqrt{0.6719}} < \frac{103 - 100}{\sqrt{0.6719}}\right)$$
$$\approx \Phi(3.6599) - \Phi(1.2200) = 0.9999 - 0.8888 = 0.1111.$$

   (c) $n = 165$, $\sigma_{\bar{Y}}^2 = \frac{\sigma_Y^2}{n} = \frac{43}{165} = 0.2606$, and

$$\Pr(\bar{Y} > 98) = 1 - \Pr(\bar{Y} \le 98) = 1 - \Pr\left(\frac{\bar{Y} - 100}{\sqrt{0.2606}} \le \frac{98 - 100}{\sqrt{0.2606}}\right)$$
$$\approx 1 - \Phi(-3.9178) = \Phi(3.9178) = 1.0000 \text{ (rounded to four decimal places).}$$

3. Denote each voter's preference by Y. $Y = 1$ if the voter prefers the incumbent and $Y = 0$ if the voter prefers the challenger. Y is a Bernoulli random variable with probability $\Pr(Y = 1) = p$ and $\Pr(Y = 0) = 1 - p$. From the solution to Exercise 3.2, Y has mean p and variance $p(1 - p)$.

   (a) $\hat{p} = \frac{215}{400} = 0.5375$.

   (b) $\widehat{\text{var}}(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n} = \frac{0.5375 \times (1 - 0.5375)}{400} = 6.2148 \times 10^{-4}$. The standard error is $SE(\hat{p}) = (\text{var}(\hat{p}))^{\frac{1}{2}} = 0.0249$.

   (c) The computed t-statistic is

$$t^{act} = \frac{\hat{p} - \mu_{p,0}}{SE(\hat{p})} = \frac{0.5375 - 0.5}{0.0249} = 1.506.$$

   Because of the large sample size $(n = 400)$, we can use Equation (3.14) in the text to get the p-value for the test $H_0: p = 0.5$ vs. $H_1: p \neq 0.5$:

$$p\text{-value} = 2\Phi(-|t^{act}|) = 2\Phi(-1.506) = 2 \times 0.066 = 0.132.$$

(d) Using Equation (3.17) in the text, the *p*-value for the test $H_0: p = 0.5$ vs. $H_1: p > 0.5$ is

$$p\text{-value} = 1 - \Phi(t^{act}) = 1 - \Phi(1.506) = 1 - 0.934 = 0.066.$$

(e) Part (c) is a two-sided test and the *p*-value is the area in the tails of the standard normal distribution outside ± (calculated *t*-statistic). Part (d) is a one-sided test and the *p*-value is the area under the standard normal distribution to the right of the calculated *t*-statistic.

(f) For the test $H_0: p = 0.5$ vs. $H_1: p > 0.5$, we cannot reject the null hypothesis at the 5% significance level. The *p*-value 0.066 is larger than 0.05. Equivalently the calculated *t*-statistic 1.506 is less than the critical value 1.645 for a one-sided test with a 5% significance level. The test suggests that the survey did not contain statistically significant evidence that the incumbent was ahead of the challenger at the time of the survey.

5. (a) (i) The size is given by $\Pr(|\hat{p} - 0.5| > .02)$, where the probability is computed assuming that $p = 0.5$.

$$\begin{aligned}\Pr(|\hat{p} - 0.5| > .02) &= 1 - \Pr(-0.02 \le \hat{p} - 0.5 \le .02)\\ &= 1 - \Pr\left(\frac{-0.02}{\sqrt{.5\times.5/1055}} \le \frac{\hat{p} - 0.5}{\sqrt{.5\times.5/1055}} \le \frac{0.02}{\sqrt{.5\times.5/1055}}\right)\\ &= 1 - \Pr\left(-1.30 \le \frac{\hat{p} - 0.5}{\sqrt{.5\times.5/1055}} \le 1.30\right)\\ &= 0.19\end{aligned}$$

where the final equality using the central limit theorem approximation

(ii) The power is given by $\Pr(|\hat{p} - 0.5| > .02)$, where the probability is computed assuming that $p = 0.53$.

$$\begin{aligned}\Pr(|\hat{p} - 0.5| > .02) &= 1 - \Pr(-0.02 \le \hat{p} - 0.5 \le .02)\\ &= 1 - \Pr\left(\frac{-0.02}{\sqrt{.53\times.47/1055}} \le \frac{\hat{p} - 0.5}{\sqrt{.53\times.47/1055}} \le \frac{0.02}{\sqrt{.53\times.47/1055}}\right)\\ &= 1 - \Pr\left(\frac{-0.05}{\sqrt{.53\times.47/1055}} \le \frac{\hat{p} - 0.53}{\sqrt{.53\times.47/1055}} \le \frac{-0.01}{\sqrt{.53\times.47/1055}}\right)\\ &= 1 - \Pr\left(-3.25 \le \frac{\hat{p} - 0.53}{\sqrt{.53\times.47/1055}} \le -0.65\right)\\ &= 0.74\end{aligned}$$

where the final equality using the central limit theorem approximation.

(b) (i) $t = \frac{0.54-0.5}{\sqrt{0.54\times0.46/1055}} = 2.61$, $\Pr(|t| > 2.61) = .01$, so that the null is rejected at the 5% level.

(ii) $\Pr(t > 2.61) = .004$, so that the null is rejected at the 5% level.

(iii) $0.54 \pm 1.96\sqrt{0.54\times0.46/1055} = 0.54 \pm 0.03$, or 0.51 to 0.57.

(iv) $0.54 \pm 2.58\sqrt{0.54\times0.46/1055} = 0.54 \pm 0.04$, or 0.50 to 0.58.

(v) $0.54 \pm 0.67\sqrt{0.54 \times 0.46/1055} = 0.54 \pm 0.01$, or 0.53 to 0.55.

(c) (i) The probability is 0.95 is any single survey, there are 20 independent surveys, so the probability if $0.95^{20} = 0.36$

(ii) 95% of the 20 confidence intervals or 19.

(d) The relevant equation is $1.96 \times \text{SE}(\hat{p}) < .01$ or $1.96 \times \sqrt{p(1-p)/n} < .01$. Thus n must be chosen so that $n > \frac{1.96^2 p(1-p)}{.01^2}$, so that the answer depends on the value of p. Note that the largest value that $p(1 - p)$ can take on is 0.25 (that is, $p = 0.5$ makes $p(1 - p)$ as large as possible). Thus if $n > \frac{1.96^2 \times 0.25}{.01^2} = 9604$, then the margin of error is less than 0.01 for all values of p.

7. The null hypothesis in that the survey is a random draw from a population with $p = 0.11$. The t-statistic is $t = \frac{\hat{p} - 0.11}{SE(\hat{p})}$, where $SE(\hat{p}) = \hat{p}(1-\hat{p})/n$. (An alternative formula for SE($\hat{p}$) is $0.11 \times (1 - 0.11)/n$, which is valid under the null hypothesis that $p = 0.11$). The value of the t-statistic is -2.71, which has a p-value of that is less than 0.01. Thus the null hypothesis $p = 0.11$ (the survey is unbiased) can be rejected at the 1% level.

9. Denote the life of a light bulb from the new process by Y. The mean of Y is μ and the standard deviation of Y is $\sigma_Y = 200$ hours. $\overline{Y}$ is the sample mean with a sample size $n = 100$. The standard deviation of the sampling distribution of $\overline{Y}$ is $\sigma_{\overline{Y}} = \frac{\sigma_Y}{\sqrt{n}} = \frac{200}{\sqrt{100}} = 20$ hours. The hypothesis test is $H_0: \mu = 2000$ vs. $H_1: \mu > 2000$. The manager will accept the alternative hypothesis if $\overline{Y} > 2100$ hours.

(a) The size of a test is the probability of erroneously rejecting a null hypothesis when it is valid. The size of the manager's test is

$$\begin{aligned}
\text{size} &= \Pr(\overline{Y} > 2100 | \mu = 2000) = 1 - \Pr(\overline{Y} \le 2100 | \mu = 2000) \\
&= 1 - \Pr\left(\frac{\overline{Y} - 2000}{20} \le \frac{2100 - 2000}{20} | \mu = 2000\right) \\
&= 1 - \Phi(5) = 1 - 0.999999713 = 2.87 \times 10^{-7}.
\end{aligned}$$

$\Pr(\overline{Y} > 2100 | \mu = 2000)$ means the probability that the sample mean is greater than 2100 hours when the new process has a mean of 2000 hours.

(b) The power of a test is the probability of correctly rejecting a null hypothesis when it is invalid. We calculate first the probability of the manager erroneously accepting the null hypothesis when it is invalid:

$$\begin{aligned}
\beta &= \Pr(\overline{Y} \le 2100 | \mu = 2150) = \Pr\left(\frac{\overline{Y} - 2150}{20} \le \frac{2100 - 2150}{20} | \mu = 2150\right) \\
&= \Phi(-2.5) = 1 - \Phi(2.5) = 1 - 0.9938 = 0.0062.
\end{aligned}$$

The power of the manager's testing is $1 - \beta = 1 - 0.0062 = 0.9938$.

(c) For a test with 5%, the rejection region for the null hypothesis contains those values of the t-statistic exceeding 1.645.

$$t^{act} = \frac{\overline{Y}^{act} - 2000}{20} > 1.645 \Rightarrow \overline{Y}^{act} > 2000 + 1.645 \times 20 = 2032.9.$$

The manager should believe the inventor's claim if the sample mean life of the new product is greater than 2032.9 hours if she wants the size of the test to be 5%.

11. Assume that n is an even number. Then $\tilde{Y}$ is constructed by applying a weight of $\frac{1}{2}$ to the $\frac{n}{2}$ "odd" observations and a weight of $\frac{3}{2}$ to the remaining $\frac{n}{2}$ observations.

$$\begin{aligned} E(\tilde{Y}) &= \frac{1}{n}\left(\frac{1}{2}E(Y_1) + \frac{3}{2}E(Y_2) + \cdots \frac{1}{2}E(Y_{n-1}) + \frac{3}{2}E(Y_n)\right) \\ &= \frac{1}{n}\left(\frac{1}{2}\cdot\frac{n}{2}\cdot\mu_Y + \frac{3}{2}\cdot\frac{n}{2}\cdot\mu_Y\right) = \mu_Y \end{aligned}$$

$$\begin{aligned} \text{var}(\tilde{Y}) &= \frac{1}{n^2}\left(\frac{1}{4}\text{var}(Y_1) + \frac{9}{4}\text{var}(Y_2) + \cdots \frac{1}{4}\text{var}(Y_{n-1}) + \frac{9}{4}\text{var}(Y_n)\right) \\ &= \frac{1}{n^2}\left(\frac{1}{4}\cdot\frac{n}{2}\cdot\sigma_Y^2 + \frac{9}{4}\cdot\frac{n}{2}\cdot\sigma_Y^2\right) = 1.25\frac{\sigma_Y^2}{n}. \end{aligned}$$

13. (a) Sample size $n = 420$, sample average $\overline{Y} = 654.2$, sample standard deviation $s_Y = 19.5$. The standard error of $\overline{Y}$ is $\text{SE}(\overline{Y}) = \frac{s_Y}{\sqrt{n}} = \frac{19.5}{\sqrt{420}} = 0.9515$. The 95% confidence interval for the mean test score in the population is

$$\mu = \overline{Y} \pm 1.96\text{SE}(\overline{Y}) = 654.2 \pm 1.96 \times 0.9515 = (652.34,\ 656.06).$$

(b) The data are: sample size for small classes $n_1 = 238$, sample average $\overline{Y}_1 = 657.4$, sample standard deviation $s_1 = 19.4$; sample size for large classes $n_2 = 182$, sample average $\overline{Y}_2 = 650.0$, sample standard deviation $s_2 = 17.9$. The standard error of $\overline{Y}_1 - \overline{Y}_2$ is $SE(\overline{Y}_1 - \overline{Y}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{19.4^2}{238} + \frac{17.9^2}{182}} = 1.8281$. The hypothesis tests for higher average scores in smaller classes is

$$H_0: \mu_1 - \mu_2 = 0 \quad \text{vs.} \quad H_1: \mu_1 - \mu_2 > 0.$$

The *t*-statistic is

$$t^{act} = \frac{\overline{Y}_1 - \overline{Y}_2}{\text{SE}(\overline{Y}_1 - \overline{Y}_2)} = \frac{657.4 - 650.0}{1.8281} = 4.0479.$$

The *p*-value for the one-sided test is:

$$p\text{-value} = 1 - \Phi(t^{act}) = 1 - \Phi(4.0479) = 1 - 0.999974147 = 2.5853 \times 10^{-5}.$$

With the small *p*-value, the null hypothesis can be rejected with a high degree of confidence. There is statistically significant evidence that the districts with smaller classes have higher average test scores.

15. Let p denote the fraction of the population that preferred Bush.

(a) $\hat{p} = 405/755 = 0.536$; $\text{SE}(\hat{p}) = .0181$; 95% confidence interval is $\hat{p} \pm 1.96\,\text{SE}(\hat{p})$ or $0.536 \pm .036$

(b) $\hat{p} = 378/756 = 0.500$; $\text{SE}(\hat{p}) = .0182$; 95% confidence interval is $\hat{p} \pm 1.96\,\text{SE}(\hat{p})$ or 0.500 ± 0.36

(c) $\hat{p}_{Sep} - \hat{p}_{Oct} = 0.036$; $\text{SE}(\hat{p}_{Sep} - \hat{p}_{Oct}) = \sqrt{\frac{0.536(1-0.536)}{755} + \frac{0.5(1-0.5)}{756}}$ (because the surveys are independent. The 95% confidence interval for the change in p is $(\hat{p}_{Sep} - \hat{p}_{Oct}) \pm 1.96\,\text{SE}(\hat{p}_{Sep} - \hat{p}_{Oct})$ or $0.036 \pm .050$. The confidence interval includes $(p_{Sep} - p_{Oct}) = 0.0$, so there is not statistically significance evidence of a change in voters' preferences.

17. (a) The 95% confidence interval is $\overline{Y}_{m,2004} - \overline{Y}_{m,1992} \pm 1.96\,\text{SE}(\overline{Y}_{m,2004} - \overline{Y}_{m,1992})$ where $\text{SE}(\overline{Y}_{m,2004} - \overline{Y}_{m,1992}) = \sqrt{\frac{S^2_{m,2004}}{n_{m,2004}} + \frac{S^2_{m,1992}}{n_{m,1992}}} = \sqrt{\frac{10.39^2}{1901} + \frac{8.70^2}{1592}} = 0.32$; the 95% confidence interval is (21.99 – 20.33) ± 0.63 or 1.66 ± 0.63.

(b) The 95% confidence interval is $\overline{Y}_{w,2004} - \overline{Y}_{w,1992} \pm 1.96\,\text{SE}(\overline{Y}_{w,2004} - \overline{Y}_{w,1992})$ where $\text{SE}(\overline{Y}_{w,2004} - \overline{Y}_{w,1992}) = \sqrt{\frac{S^2_{w,2004}}{n_{w,2004}} + \frac{S^2_{w,1992}}{n_{w,1992}}} = \sqrt{\frac{8.16^2}{1739} + \frac{6.90^2}{1370}} = 0.27$; the 95% confidence interval is $(18.47 - 17.60) \pm 0.53$ or 0.87 ± 0.53.

(c) The 95% confidence interval is $(\overline{Y}_{m,2004} - \overline{Y}_{m,1992}) - (\overline{Y}_{w,2004} - \overline{Y}_{w,1992}) \pm 1.96\,\text{SE}[(\overline{Y}_{m,2004} - \overline{Y}_{m,1992}) - (\overline{Y}_{w,2004} - \overline{Y}_{w,1992})]$, where $\text{SE}[(\overline{Y}_{m,2004} - \overline{Y}_{m,1992}) - (\overline{Y}_{w,2004} - \overline{Y}_{w,1992})] = \sqrt{\frac{S^2_{m,2004}}{n_{m,2004}} + \frac{S^2_{m,1992}}{n_{m,1992}} + \frac{S^2_{w,2004}}{n_{w,2004}} + \frac{S^2_{w,2004}}{n_{w,2004}}} = \sqrt{\frac{10.39^2}{1901} + \frac{8.70^2}{1592} + \frac{8.16^2}{1739} + \frac{6.90^2}{1370}} = 0.42.$ The 95% confidence interval is (21.99 – 20.33) – (18.47 – 17.60) ± 1.96 × 0.42 or 0.79 ± 0.82.

19. (a) No. $E(Y_i^2) = \sigma_Y^2 + \mu_Y^2$ and $E(Y_iY_j) = \mu_Y^2$ for $i \neq j$. Thus

$$E(\overline{Y}^2) = E\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right)^2 = \frac{1}{n^2}\sum_{i=1}^{n} E(Y_i^2) + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i} E(Y_iY_j)$$
$$= \mu_Y^2 + \frac{1}{n}\sigma_Y^2$$

(b) Yes. If $\overline{Y}$ gets arbitrarily close to μ_Y with probability approaching 1 as n gets large, then $\overline{Y}^2$ gets arbitrarily close to μ_Y^2 with probability approaching 1 as n gets large. (As it turns out, this is an example of the "continuous mapping theorem" discussed in Chapter 17.)

21. Set $n_m = n_w = n$, and use equation (3.19) write the squared SE of $\overline{Y}_m - \overline{Y}_w$ as

$$[SE(\overline{Y}_m - \overline{Y}_w)]^2 = \frac{\frac{1}{(n-1)}\sum_{i=1}^{n}(Y_{mi} - \overline{Y}_m)^2}{n} + \frac{\frac{1}{(n-1)}\sum_{i=1}^{n}(Y_{wi} - \overline{Y}_w)^2}{n}$$

$$= \frac{\sum_{i=1}^{n}(Y_{mi} - \overline{Y}_m)^2 + \sum_{i=1}^{n}(Y_{wi} - \overline{Y}_w)^2}{n(n-1)}.$$

Similary, using equation (3.23)

$$[SE_{pooled}(\overline{Y}_m - \overline{Y}_w)]^2 = \frac{\frac{1}{2(n-1)}\left[\sum_{i=1}^{n}(Y_{mi} - \overline{Y}_m)^2 + \frac{1}{(n-1)}\sum_{i=1}^{n}(Y_{wi} - \overline{Y}_w)^2\right]}{2n}$$

$$= \frac{\sum_{i=1}^{n}(Y_{mi} - \overline{Y}_m)^2 + \sum_{i=1}^{n}(Y_{wi} - \overline{Y}_w)^2}{n(n-1)}.$$

Chapter 4
Linear Regression with One Regressor

■ Solutions to Exercises

1. (a) The predicted average test score is

$$\widehat{TestScore} = 520.4 - 5.82 \times 22 = 392.36$$

(b) The predicted change in the classroom average test score is

$$\Delta\widehat{TestScore} = (-5.82 \times 19) - (-5.82 \times 23) = 23.28$$

(c) Using the formula for $\hat{\beta}_0$ in Equation (4.8), we know the sample average of the test scores across the 100 classrooms is

$$\overline{TestScore} = \hat{\beta}_0 + \hat{\beta}_1 \times \overline{CS} = 520.4 - 5.82 \times 21.4 = 395.85.$$

(d) Use the formula for the standard error of the regression (SER) in Equation (4.19) to get the sum of squared residuals:

$$SSR = (n-2)SER^2 = (100-2) \times 11.5^2 = 12961.$$

Use the formula for R^2 in Equation (4.16) to get the total sum of squares:

$$TSS = \frac{SSR}{1-R^2} = \frac{12961}{1-0.08^2} = 13044.$$

The sample variance is $s_Y^2 = \frac{TSS}{n-1} = \frac{13044}{99} = 131.8.$ Thus, standard deviation is $s_Y = \sqrt{s_Y^2} = 11.5.$

3. (a) The coefficient 9.6 shows the marginal effect of *Age* on *AWE*; that is, *AWE* is expected to increase by $9.6 for each additional year of age. 696.7 is the intercept of the regression line. It determines the overall level of the line.

(b) *SER* is in the same units as the dependent variable (*Y*, or *AWE* in this example). Thus *SER* is measures in dollars per week.

(c) R^2 is unit free.

(d) (i) $696.7 + 9.6 \times 25 = \$936.7$;

(ii) $696.7 + 9.6 \times 45 = \$1,128.7$

(e) No. The oldest worker in the sample is 65 years old. 99 years is far outside the range of the sample data.

(f) No. The distribution of earning is positively skewed and has kurtosis larger than the normal.

(g) $\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1\overline{X},$ so that $\overline{Y} = \hat{\beta}_0 + \hat{\beta}_1\overline{X}.$ Thus the sample mean of *AWE* is 696.7 + 9.6 × 41.6 = $1,096.06.

5. (a) u_i represents factors other than time that influence the student's performance on the exam including amount of time studying, aptitude for the material, and so forth. Some students will have studied more than average, other less; some students will have higher than average aptitude for the subject, others lower, and so forth.

   (b) Because of random assignment u_i is independent of X_i. Since u_i represents deviations from average $E(u_i) = 0$. Because u and X are independent $E(u_i|X_i) = E(u_i) = 0$.

   (c) (2) is satisfied if this year's class is typical of other classes, that is, students in this year's class can be viewed as random draws from the population of students that enroll in the class. (3) is satisfied because $0 \leq Y_i \leq 100$ and X_i can take on only two values (90 and 120).

   (d) (i) $49 + 0.24 \times 90 = 70.6$; $49 + 0.24 \times 120 = 77.8$; $49 + 0.24 \times 150 = 85.0$

   (ii) $0.24 \times 10 = 2.4$.

7. The expectation of $\hat{\beta}_0$ is obtained by taking expectations of both sides of Equation (4.8):

$$E(\hat{\beta}_0) = E(\overline{Y} - \hat{\beta}_1\overline{X}) = E\left[\left(\beta_0 + \beta_1\overline{X} + \frac{1}{n}\sum_{i=1}^{n} u_i\right) - \hat{\beta}_1\overline{X}\right]$$
$$= \beta_0 + E(\beta_1 - \hat{\beta}_1)\overline{X} + \frac{1}{n}\sum_{i=1}^{n} E(u_i|X_i) = \beta_0,$$

where the third equality in the above equation has used the facts that $\hat{\beta}_1$ is unbiased so $E(\beta_1 - \hat{\beta}_1) = 0$ and $E(u_i|X_i) = 0$.

9. (a) With $\hat{\beta}_1 = 0$, $\hat{\beta}_0 = \overline{Y}$, and $\hat{Y}_i = \hat{\beta}_0 = \overline{Y}$. Thus ESS = 0 and $R^2 = 0$.

   (b) If $R^2 = 0$, then ESS = 0, so that $\hat{Y}_i = \overline{Y}$ for all i. But $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$, so that $\hat{Y}_i = \overline{Y}$ for all i, which implies that $\hat{\beta}_1 = 0$, or that X_i is constant for all i. If X_i is constant for all i, then $\sum_{i-1}^{n}(X_i - \overline{X})^2 = 0$ and $\hat{\beta}_1$ is undefined (see equation (4.7)).

11. (a) The least squares objective function is $\sum_{i=1}^{n}(Y_i - b_1X_i)^2$. Differentiating with respect to b_1 yields $\frac{\partial \sum_{i=1}^{n}(Y_i - b_1X_i)^2}{\partial b_1} = -2\sum_{i=1}^{n} X_i(Y_i - b_1X_i)$. Setting this zero, and solving for the least squares estimator yields $\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_iY_i}{\sum_{i=1}^{n} X_i^2}$.

   (b) Following the same steps in (a) yields $\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i(Y_i - 4)}{\sum_{i=1}^{n} X_i^2}$

Chapter 5
Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals

■ Solutions to Exercises

1 (a) The 95% confidence interval for β_1 is $\{-5.82 \pm 1.96 \times 2.21\}$, that is $-10.152 \le \beta_1 \le -1.4884$.

(b) Calculate the *t*-statistic:

$$t^{act} = \frac{\hat{\beta}_1 - 0}{\mathrm{SE}(\hat{\beta}_1)} = \frac{-5.82}{2.21} = -2.6335.$$

The *p*-value for the test $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$ is

$$p\text{-value} = 2\Phi(-|t^{act}|) = 2\Phi(-2.6335) = 2 \times 0.0042 = 0.0084.$$

The *p*-value is less than 0.01, so we can reject the null hypothesis at the 5% significance level, and also at the 1% significance level.

(c) The *t*-statistic is

$$t^{act} = \frac{\hat{\beta}_1 - (-5.6)}{\mathrm{SE}(\hat{\beta}_1)} = \frac{0.22}{2.21} = 0.10$$

The *p*-value for the test $H_0: \beta_1 = -5.6$ vs. $H_1: \beta_1 \neq -5.6$ is

$$p\text{-value} = 2\Phi(-|t^{act}|) = 2\Phi(-0.10) = 0.92$$

The *p*-value is larger than 0.10, so we cannot reject the null hypothesis at the 10%, 5% or 1% significance level. Because $\beta_1 = -5.6$ is not rejected at the 5% level, this value is contained in the 95% confidence interval.

(d) The 99% confidence interval for β_0 is $\{520.4 \pm 2.58 \times 20.4\}$, that is, $467.7 \le \beta_0 \le 573.0$.

3. The 99% confidence interval is 1.5 × {3.94 ± 2.58 × 0.31) or 4.71 lbs ≤ WeightGain ≤ 7.11 lbs.

5 (a) The estimated gain from being in a small class is 13.9 points. This is equal to approximately 1/5 of the standard deviation in test scores, a moderate increase.

(b) The *t*-statistic is $t^{act} = \frac{13.9}{2.5} = 5.56$, which has a *p*-value of 0.00. Thus the null hypothesis is rejected at the 5% (and 1%) level.

(c) 13.9 ± 2.58 × 2.5 = 13.9 ± 6.45.

7. (a) The *t*-statistic is $\frac{3.2}{1.5} = 2.13$ with a *p*-value of 0.03; since the *p*-value is less than 0.05, the null hypothesis is rejected at the 5% level.

(b) 3.2 ± 1.96 × 1.5 = 3.2 ± 2.94

(c) Yes. If Y and X are independent, then $\beta_1 = 0$; but this null hypothesis was rejected at the 5% level in part (a).

(d) β_1 would be rejected at the 5% level in 5% of the samples; 95% of the confidence intervals would contain the value $\beta_1 = 0$.

9. (a) $\bar{\beta} = \frac{1}{\bar{X}}\frac{1}{n}(Y_1 + Y_2 + \cdots + Y_n)$ so that it is linear function of $Y_1, Y_2, \ldots, Y_n$.

(b) $E(Y_i|X_1, \ldots, X_n) = \beta_1 X_i$, thus

$$E(\bar{\beta}|X_1,\ldots,X_n) = E\frac{1}{\bar{X}}\frac{1}{n}(Y_1 + Y_2 + \cdots + Y_n)|X_1,\ldots,X_n)$$
$$= \frac{1}{\bar{X}}\frac{1}{n}\beta_1(X_1 + \cdots + X_n) = \beta_1$$

11. Using the results from 5.10, $\hat{\beta}_0 = \bar{Y}_m$ and $\hat{\beta}_1 = \bar{Y}_w - \bar{Y}_m$. From Chapter 3, $\text{SE}(\bar{Y}_m) = \frac{S_m}{\sqrt{n_m}}$ and $\text{SE}(\bar{Y}_w - \bar{Y}_m) = \sqrt{\frac{s_m^2}{n_m} + \frac{s_w^2}{n_w}}$. Plugging in the numbers $\hat{\beta}_0 = 523.1$ and $\text{SE}(\hat{\beta}_0) = 6.22$; $\hat{\beta}_1 = -38.0$ and $\text{SE}(\hat{\beta}_1) = 7.65$.

13. (a) Yes

(b) Yes

(c) They would be unchanged

(d) (a) is unchanged; (b) is no longer true as the errors are not conditionally homosckesdastic.

15. Because the samples are independent, $\hat{\beta}_{m,1}$ and $\hat{\beta}_{w,1}$ are independent. Thus $\text{var}(\hat{\beta}_{m,1} - \hat{\beta}_{w,1}) = \text{var}(\hat{\beta}_{m,1}) + \text{var}(\hat{\beta}_{w,1})$. $\text{Var}(\hat{\beta}_{m,1})$ is consistently estimated as $[SE(\hat{\beta}_{m,1})]^2$ and $\text{Var}(\hat{\beta}_{w,1})$ is consistently estimated as $[SE(\hat{\beta}_{w,1})]^2$, so that $\text{var}(\hat{\beta}_{m,1} - \hat{\beta}_{w,1})$ is consistently estimated by $[SE(\hat{\beta}_{m,1})]^2 + [SE(\hat{\beta}_{w,1})]^2$, and the result follows by noting the SE is the square root of the estimated variance.

Chapter 6
Linear Regression with Multiple Regressors

■ Solutions to Exercises

1. By equation (6.15) in the text, we know

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1-R^2).$$

Thus, that values of $\bar{R}^2$ are 0.175, 0.189, and 0.193 for columns (1)–(3).

3. (a) On average, a worker earns \$0.29/hour more for each year he ages.
   (b) Sally's earnings prediction is $4.40 + 5.48 \times 1 - 2.62 \times 1 + 0.29 \times 29 = 15.67$ dollars per hour. Betsy's earnings prediction is $4.40 + 5.48 \times 1 - 2.62 \times 1 + 0.29 \times 34 = 17.12$ dollars per hour. The difference is 1.45

5. (a) \$23,400 (recall that *Price* is measured in \$1000s).
   (b) In this case $\Delta BDR = 1$ and $\Delta Hsize = 100$. The resulting expected change in price is $23.4 + 0.156 \times 100 = 39.0$ thousand dollars or \$39,000.
   (c) The loss is \$48,800.
   (d) From the text $\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1-R^2)$, so $R^2 = 1 - \frac{n-k-1}{n-1}(1-\bar{R}^2)$, thus, $R^2 = 0.727$.

7. (a) The proposed research in assessing the presence of gender bias in setting wages is too limited. There might be some potentially important determinants of salaries: type of engineer, amount of work experience of the employee, and education level. The gender with the lower wages could reflect the type of engineer among the gender, the amount of work experience of the employee, or the education level of the employee. The research plan could be improved with the collection of additional data as indicated and an appropriate statistical technique for analyzing the data would be a multiple regression in which the dependent variable is wages and the independent variables would include a dummy variable for gender, dummy variables for type of engineer, work experience (time units), and education level (highest grade level completed). The potential importance of the suggested omitted variables makes a "difference in means" test inappropriate for assessing the presence of gender bias in setting wages.
   (b) The description suggests that the research goes a long way towards controlling for potential omitted variable bias. Yet, there still may be problems. Omitted from the analysis are characteristics associated with behavior that led to incarceration (excessive drug or alcohol use, gang activity, and so forth), that might be correlated with future earnings. Ideally, data on these variables should be included in the analysis as additional control variables.

9. For omitted variable bias to occur, two conditions must be true: X_1 (the included regressor) is correlated with the omitted variable, and the omitted variable is a determinant of the dependent variable. Since X_1 and X_2 are uncorrelated, the estimator of β_1 does not suffer from omitted variable bias.

11. (a)

$$\sum (Y_i - b_1 X_{1i} - b_2 X_{2i})^2$$

(b)

$$\frac{\partial \sum (Y_i - b_1 X_{1i} - b_2 X_{2i})^2}{\partial b_1} = -2\sum X_{1i}(Y_i - b_1 X_{1i} - b_2 X_{2i})$$

$$\frac{\partial \sum (Y_i - b_1 X_{1i} - b_2 X_{2i})^2}{\partial b_2} = -2\sum X_{2i}(Y_i - b_1 X_{1i} - b_2 X_{2i})$$

(c) From (b), $\hat{\beta}_1$ satisfies

$$\sum X_{1i}(Y_i - \hat{\beta}_1 X_{1i} - \hat{\beta}_1 X_{2i}) = 0$$

or

$$\hat{\beta}_1 = \frac{\sum X_{1i} Y_i - \hat{\beta}_2 \sum X_{1i} X_{2i}}{\sum X_{1i}^2}$$

and the result follows immediately.

(d) Following analysis as in (c)

$$\hat{\beta}_2 = \frac{\sum X_{2i} Y_i - \hat{\beta}_1 \sum X_{1i} X_{2i}}{\sum X_{2i}^2}$$

and substituting this into the expression for $\hat{\beta}_1$ in (c) yields

$$\hat{\beta}_1 = \frac{\sum X_{1i} Y \frac{\sum X_{2i} Y_i - \hat{\beta}_1 \sum X_{1i} X_{2i}}{\sum X_{2i}^2} \sum X_{1i} X_{2i}}{\sum X_{1i}^2}.$$

Solving for $\hat{\beta}_1$ yields:

$$\hat{\beta}_1 = \frac{\sum X_{2i}^2 \sum X_{1i} Y_i - \sum X_{1i} X_{2i} \sum X_{2i} Y_i}{\sum X_{1i}^2 \sum X_{2i}^2 - (\sum X_{1i} X_{2i})^2}$$

(e) The least squares objective function is $\sum (Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i})^2$ and the partial derivative with respect to b_0 is

$$\frac{\partial \sum (Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i})^2}{\partial b_0} = -2\sum (Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i}).$$

Setting this to zero and solving for $\hat{\beta}_0$ yields: $\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}_1 - \hat{\beta}_2 \overline{X}_2$.

(f)

$$\frac{\partial}{\partial \beta_1} = 0 \Rightarrow -\sum X_{1i}Y_i + \hat{\beta}_2 \sum X_{1i}Y_{2i} + \hat{\beta}_1 \sum X_{1i}^2 = 0$$

$$\hat{\beta}_1 \sum X_{1i}^2 = \sum X_{1i}Y_i - \hat{\beta}_2 \sum X_{1i}X_{2i}$$

$$\hat{\beta}_1 = \frac{\sum X_{1i}Y_i - \beta_2 \sum X_{1i}Y_{2i}}{\sum X_{1i}^2}$$

$$\frac{\partial 2}{\partial \beta_2} = 0 \Rightarrow -\sum X_{2i}Y_i + \hat{\beta}_1 \sum X_{1i}X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 = 0$$

$$\hat{\beta}_2 \sum X_{2i}^2 = \sum X_{2i}Y_i - \hat{\beta}_1 \sum X_{1i}X_{2i}$$

$$\hat{\beta}_2 = \frac{\sum X_{2i}Y_i - \beta_1 \sum X_{1i}Y_{2i}}{\sum X_{2i}^2}$$

(g)

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u$$

$$u_i = Y_i - (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i})$$

$$\sum_{i=1}^{n} u_i^2 = \sum [Y_i - (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i})]^2$$

$$= \sum [(Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i})(Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i})]$$

$$= \sum [Y_i^2 - Y_i\beta_0 - Y_i\beta_1 X_{1i} - Y_i\beta_2 X_{2i} - \beta_0 Y_i + \beta_0^2 + \beta_0\beta_1 X_{1i} + \beta_0\beta_2 X_{2i}$$

$$- \beta_1 X_{1i}Y_i + \beta_1 X_{1i}\beta_0 + \beta_1^2 X_{1i}^2 + \beta_1 X_{1i}\beta_2 X_{2i} - \beta_2 X_{2i}Y_i + \beta_2 X_{2i}\beta_0 + \beta_2 X_{2i}\beta_1 X_{1i} + \beta_2^2 X_{2i}^2]$$

$$= \sum [Y_i^2 - 2\beta_0 Y_i - 2\beta_1 X_{1i}Y_i - 2\beta_2 X_{2i}Y_i + \beta_0^2 + 2\beta_0\beta_1 X_{1i} + 2\beta_0\beta_2 X_{2i}$$

$$+ 2\beta_1\beta_2 X_{1i}X_{2i} + \beta_1^2 X_{1i}^2 + \beta_2^2 X_{2i}^2]$$

$$= \frac{1}{n}\sum Y_i^2 - 2\beta_0 \bar{Y} - 2\beta_1 \sum X_{1i}Y_i - 2\beta_2 \sum X_{2i}Y_i + \beta_0^2 + 2\beta_0\beta_1 \bar{X}_1 + 2\beta_0\beta_2 \bar{X}_2$$

$$+ 2\beta_1\beta_2 \sum X_{1i}X_{2i} + \beta_1^2 \sum X_{1i}^2 + \beta_2^2 \sum X_{2i}^2$$

$$\frac{\partial}{\partial \beta_0} = -2\bar{Y} + 2\hat{\beta}_0 + 2\hat{\beta}_1 \bar{X}_1 + 2\hat{\beta}_2 \bar{X}_2 = 0 \Rightarrow 2\hat{\beta}_0 = 2\bar{Y} - 2\hat{\beta}_1 \bar{X}_1 - 2\hat{\beta}_2 \bar{X}_2$$

$$\Rightarrow \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$$

Chapter 7
Hypothesis Tests and Confidence Intervals in Multiple Regression

■ Solutions to Exercises

1.

Regressor	**(1)**	**(2)**	**(3)**
College (X_1)	5.46** (0.21)	5.48** (0.21)	5.44** (0.21)
Female (X_2)	–2.64** (0.20)	–2.62** (0.20)	–2.62** (0.20)
Age (X_3)		0.29** (0.04)	0.29** (0.04)
Ntheast (X_4)			0.69* (0.30)
Midwest (X_5)			0.60* (0.28)
South (X_6)			–0.27 (0.26)
Intercept	12.69** (0.14)	4.40** (1.05)	3.75** (1.06)

(a) The *t*-statistic is $5.46/0.21 = 26.0 > 1.96$, so the coefficient is statistically significant at the 5% level. The 95% confidence interval is $5.46 \pm 1.96 \times 0.21$.

(b) *t*-statistic is $-2.64/0.20 = -13.2$, and $13.2 > 1.96$, so the coefficient is statistically significant at the 5% level. The 95% confidence interval is $-2.64 \pm 1.96 \times 0.20$.

3. (a) Yes, age is an important determinant of earnings. Using a *t*-test, the *t*-statistic is $\frac{0.29}{0.04} = 7.25$, with a *p*-value of 4.2×10^{-13}, implying that the coefficient on age is statistically significant at the 1% level. The 95% confidence interval is $0.29 \pm 1.96 \times 0.04$.

(b) $\Delta Age \times [0.29 \pm 1.96 \times 0.04] = 5 \times [0.29 \pm 1.96 \times 0.04] = 1.45 \pm 1.96 \times 0.20 =$ \$1.06 to \$1.84

5. The *t*-statistic for the difference in the college coefficients is $t = (\hat{\beta}_{college,1998} - \hat{\beta}_{college,1992})/SE(\hat{\beta}_{college,1998} - \hat{\beta}_{college,1992})$. Because $\hat{\beta}_{college,1998}$ and $\hat{\beta}_{college,1992}$ are computed from independent samples, they are independent, which means that $\text{cov}(\hat{\beta}_{college,1998}, \hat{\beta}_{college,1992}) = 0$ Thus, $\text{var}(\hat{\beta}_{college,1998} - \hat{\beta}_{college,1992}) = \text{var}(\hat{\beta}_{college,1998}) + \text{var}(\hat{\beta}_{college,1998})$. This implies that $SE(\hat{\beta}_{college,1998} - \hat{\beta}_{college,1992}) = (0.21^2 + 0.20^2)^{\frac{1}{2}}$. Thus, $t^{act} = \frac{5.48-5.29}{(0.21^2+0.20^2)^{\frac{1}{2}}} = 0.6552$. There is no significant change since the calculated *t*-statistic is less than 1.96, the 5% critical value.

7. (a) The *t*-statistic is $\frac{0.485}{2.61} = 0.186 < 1.96$. Therefore, the coefficient on BDR is not statistically significantly different from zero.

   (b) The coefficient on *BDR* measures the *partial effect* of the number of bedrooms holding house size (*Hsize*) constant. Yet, the typical 5-bedroom house is much larger than the typical 2-bedroom house. Thus, the results in (a) says little about the conventional wisdom.

   (c) The 99% confidence interval for effect of lot size on price is 2000 × [.002 ± 2.58 × .00048] or 1.52 to 6.48 (in thousands of dollars).

   (d) Choosing the scale of the variables should be done to make the regression results easy to read and to interpret. If the lot size were measured in thousands of square feet, the estimate coefficient would be 2 instead of 0.002.

   (e) The 10% critical value from the $F_{2,\infty}$ distribution is 2.30. Because 0.08 < 2.30, the coefficients are not jointly significant at the 10% level.

9. (a) Estimate

$$Y_i = \beta_0 + \gamma X_{1i} + \beta_2 (X_{1i} + X_{2i}) + u_i$$

   and test whether $\gamma = 0$.

   (b) Estimate

$$Y_i = \beta_0 + \gamma X_{1i} + \beta_2 (X_{2i} - aX_{1i}) + u_i$$

   and test whether $\gamma = 0$.

   (c) Estimate

$$Y_i - X_{1i} = \beta_0 + \gamma X_{1i} + \beta_2 (X_{2i} - X_{1i}) + u_i$$

   and test whether $\gamma = 0$.

Chapter 8
Nonlinear Regression Functions

■ Solutions to Exercises

1. (a) The percentage increase in sales is $100 \times \frac{198-196}{196} = 1.0204\%.$ The approximation is 100 × [ln (198) – ln (196)] = 1.0152%.

   (b) When $Sales_{2002}$ = 205, the percentage increase is $100 \times \frac{205-196}{196} = 4.5918\%$ and the approximation is 100 × [ln (205) – ln (196)] = 4.4895%. When $Sales_{2002}$ = 250, the percentage increase is $100 \times \frac{250-196}{196} = 27.551\%$ and the approximation is 100 × [ln (250) – ln (196)] = 24.335%. When $Sales_{2002}$ = 500, the percentage increase is $100 \times \frac{500-196}{196} = 155.1\%$ and the approximation is 100 × [ln (500) – ln (196)] = 93.649%.

   (c) The approximation works well when the change is small. The quality of the approximation deteriorates as the percentage change increases.

3 (a) The regression functions for hypothetical values of the regression coefficients that are consistent with the educator's statement are: $\beta_1 > 0$ and $\beta_2 < 0$. When $TestScore$ is plotted against STR the regression will show three horizontal segments. The first segment will be for values of $STR < 20;$ the next segment for $20 \le STR \le 25;$ the final segment for $STR > 25.$ The first segment will be higher than the second, and the second segment will be higher than the third.

   (b) It happens because of perfect multicollinearity. With all three class size binary variables included in the regression, it is impossible to compute the OLS estimates because the intercept is a perfect linear function of the three class size regressors.

5. (a) (1) The demand for older journals is less elastic than for younger journals because the interaction term between the log of journal age and price per citation is positive. (2) There is a linear relationship between log price and log of quantity follows because the estimated coefficients on log price squared and log price cubed are both insignificant. (3) The demand is greater for journals with more characters follows from the positive and statistically significant coefficient estimate on the log of characters.

   (b) (i) The effect of ln(*Price per citation*) is given by [–0.899 + 0.141 × ln(*Age*)] × ln(*Price per citation*). Using *Age* = 80, the elasticity is [–0.899 + 0.141 × ln(80)] = –0.28.

   (ii) As described in equation (8.8) and the footnote on page 263, the standard error can be found by dividing 0.28, the absolute value of the estimate, by the square root of the *F*-statistic testing $\beta_{ln(Price\ per\ citation)} + \ln(80) \times \beta_{ln(Age) \times ln(Price\ per\ citation)} = 0.$

   (c) $\ln\left(\frac{Characters}{a}\right) = \ln(Characters) - \ln(a)$ for any constant *a*. Thus, estimated parameter on *Characters* will not change and the constant (intercept) will change.

7. (a) (i) ln(*Earnings*) for females are, on average, 0.44 lower for men than for women.

   (ii) The error term has a standard deviation of 2.65 (measured in log-points).

(iii) Yes. But the regression does not control for many factors (size of firm, industry, profitability, experience and so forth).

(iv) No. In isolation, these results do imply gender discrimination. Gender discrimination means that two workers, identical in every way but gender, are paid different wages. Thus, it is also important to control for characteristics of the workers that may affect their productivity (education, years of experience, etc.) If these characteristics are systematically different between men and women, then they may be responsible for the difference in mean wages. (If this were true, it would raise an interesting and important question of why women tend to have less education or less experience than men, but that is a question about something other than gender discrimination.) These are potentially important omitted variables in the regression that will lead to bias in the OLS coefficient estimator for *Female*. Since these characteristics were not controlled for in the statistical analysis, it is premature to reach a conclusion about gender discrimination.

(b) (i) If *MarketValue* increases by 1%, earnings increase by 0.37%

(ii) *Female* is correlated with the two new included variables and at least one of the variables is important for explaining ln(*Earnings*). Thus the regression in part (a) suffered from omitted variable bias.

(c) Forgetting about the effect or *Return*, whose effects seems small and statistically insignificant, the omitted variable bias formula (see equation (6.1)) suggests that *Female* is negatively correlated with ln(*MarketValue*).

9. Note that

$$\begin{aligned} Y &= \beta_0 + \beta_1 X + \beta_2 X^2 \\ &= \beta_0 + (\beta_1 + 21\beta_2)X + \beta_2(X^2 - 21X). \end{aligned}$$

Define a new independent variable $Z = X^2 - 21X$, and estimate

$$Y = \beta_0 + \gamma X + \beta_2 Z + u_i.$$

The confidence interval is $\hat{\gamma} \pm 1.96 \times \mathrm{SE}\left(\hat{\gamma}\right)$.

Chapter 9
Assessing Studies Based on Multiple Regression

■ Solutions to Exercises

1. As explained in the text, potential threats to external validity arise from differences between the population and setting studied and the population and setting of interest. The statistical results based on New York in the 1970's are likely to apply to Boston in the 1970's but not to Los Angeles in the 1970's. In 1970, New York and Boston had large and widely used public transportation systems. Attitudes about smoking were roughly the same in New York and Boston in the 1970s. In contrast, Los Angeles had a considerably smaller public transportation system in 1970. Most residents of Los Angeles relied on their cars to commute to work, school, and so forth.

   The results from New York in the 1970's are unlikely to apply to New York in 2002. Attitudes towards smoking changed significantly from 1970 to 2002.

3. The key is that the selected sample contains only employed women. Consider two women, Beth and Julie. Beth has no children; Julie has one child. Beth and Julie are otherwise identical. Both can earn $25,000 per year in the labor market. Each must compare the $25,000 benefit to the costs of working. For Beth, the cost of working is forgone leisure. For Julie, it is forgone leisure and the costs (pecuniary and other) of child care. If Beth is just on the margin between working in the labor market or not, then Julie, who has a higher opportunity cost, will decide not to work in the labor market. Instead, Julie will work in "home production," caring for children, and so forth. Thus, on average, women with children who decide to work are women who earn higher wages in the labor market.

5 (a)

$$Q = \frac{\gamma_1\beta_0 - \gamma_0\beta_1}{\gamma_1 - \beta_1} + \frac{\gamma_1 u - \beta_1 v}{\gamma_1 - \beta_1}.$$

and

$$P = \frac{\beta_0 - \gamma_0}{\gamma_1 - \beta_1} + \frac{u - v}{\gamma_1 - \beta_1}.$$

(b)

$$E(Q) = \frac{\gamma_1\beta_0 - \gamma_0\beta_1}{\gamma_1 - \beta_1}, \quad E(P) = \frac{\beta_0 - \gamma_0}{\gamma_1 - \beta_1}$$

(c)

$$Var(Q) = \left(\frac{1}{\gamma_1 - \beta_1}\right)^2 (\gamma_1^2\sigma_u^2 + \beta_1^2\sigma_v^2), \quad Var(P) = \left(\frac{1}{\gamma_1 - \beta_1}\right)^2 (\sigma_u^2 + \sigma_v^2), \text{ and}$$

$$Cov(P, Q) = \left(\frac{1}{\gamma_1 - \beta_1}\right)^2 (\gamma_1\sigma_u^2 + \beta_1\sigma_V^2)$$

(d) (i)

$$\hat{\beta}_1 \xrightarrow{p} \frac{Cov(Q, P)}{Var(P)} = \frac{\gamma_1\sigma_u^2 + \beta_1\sigma_V^2}{\sigma_u^2 + \sigma_V^2}, \quad \hat{\beta}_0 \xrightarrow{p} E(Q) - E(P)\frac{Cov(P, Q)}{Var(P)}$$

(ii) $\hat{\beta}_1 - \beta_1 \xrightarrow{p} \frac{\sigma_u^2(\gamma_1 - \beta_1)}{\sigma_u^2 + \sigma_V^2} > 0,$ using the fact that $\gamma_1 > 0$ (supply curves slope up) and $\beta_1 < 0$ (demand curves slope down).

7. (a) True. Correlation between regressors and error terms means that the OLS estimator is inconsistent.
   (b) True.

9. Both regressions suffer from omitted variable bias so that they will not provide reliable estimates of the causal effect of income on test scores. However, the nonlinear regression in (8.18) fits the data well, so that it could be used for forecasting.

11. Again, there are reasons for concern. Here are a few.

    Internal consistency: To the extent that price is affected by demand, there may be simultaneous equation bias.

    External consistency: The internet and introduction of "*E*-journals" may induce important changes in the market for academic journals so that the results for 2000 may not be relevant in 2008.

Chapter 10
Regression with Panel Data

■ Solutions to Exercises

1. (a) With a \$1 increase in the beer tax, the expected number of lives that would be saved is 0.45 per 10,000 people. Since New Jersey has a population of 8.1 million, the expected number of lives saved is 0.45 × 810 = 364.5. The 95% confidence interval is (0.45 ± 1.96 × 0.22) × 810 = [15.228, 713.77].

   (b) When New Jersey lowers its drinking age from 21 to 18, the expected fatality rate increases by 0.028 deaths per 10,000. The 95% confidence interval for the change in death rate is 0.028 ± 1.96× 0.066 = [–0.1014, 0.1574]. With a population of 8.1 million, the number of fatalities will increase by 0.028 × 810 = 22.68 with a 95% confidence interval [–0.1014, 0.1574] × 810 = [–82.134, 127.49].

   (c) When real income per capita in new Jersey increases by 1%, the expected fatality rate increases by 1.81 deaths per 10,000. The 90% confidence interval for the change in death rate is 1.81 ± 1.64 × 0.47 = [1.04, 2.58]. With a population of 8.1 million, the number of fatalities will increase by 1.81 × 810 = 1466.1 with a 90% confidence interval [1.04, 2.58] × 810 = [840, 2092].

   (d) The low *p*-value (or high *F*-statistic) associated with the *F*-test on the assumption that time effects are zero suggests that the time effects should be included in the regression.

   (e) The difference in the significance levels arises primarily because the estimated coefficient is higher in (5) than in (4). However, (5) leaves out two variables (unemployment rate and real income per capita) that are statistically significant. Thus, the estimated coefficient on *Beer Tax* in (5) may suffer from omitted variable bias. The results from (4) seem more reliable. In general, statistical significance should be used to measure reliability only if the regression is well-specified (no important omitted variable bias, correct functional form, no simultaneous causality or selection bias, and so forth.)

   (f) Define a binary variable *west* which equals 1 for the western states and 0 for the other states. Include the interaction term between the binary variable *west* and the unemployment rate, *west* × (unemployment rate), in the regression equation corresponding to column (4). Suppose the coefficient associated with unemployment rate is β, and the coefficient associated with west × (unemployment rate) is γ. Then β captures the effect of the unemployment rate in the eastern states, and $\beta + \gamma$ captures the effect of the unemployment rate in the western states. The difference in the effect of the unemployment rate in the western and eastern states is γ. Using the coefficient estimate $(\hat{\gamma})$ and the standard error $\text{SE}(\hat{\gamma})$, you can calculate the *t*-statistic to test whether γ is statistically significant at a given significance level.

3. The five potential threats to the internal validity of a regression study are: omitted variables, misspecification of the functional form, imprecise measurement of the independent variables, sample selection, and simultaneous causality. You should think about these threats one-by-one. Are there important omitted variables that affect traffic fatalities and that may be correlated with the other variables included in the regression? The most obvious candidates are the safety of roads, weather, and so forth. These variables are essentially constant over the sample period, so their effect is captured by the state fixed effects. You may think of something that we missed. Since most of the variables are binary variables, the largest functional form choice involves the *Beer Tax* variable. A linear specification is used in the text, which seems generally consistent with the data in Figure 8.2. To check the reliability of the linear specification, it would be useful to consider a log specification or a quadratic. Measurement error does not appear to a problem, as variables like traffic fatalities and taxes are accurately measured. Similarly, sample selection is a not a problem because data were used from all of the states. Simultaneous causality could be a potential problem. That is, states with high fatality rates might decide to increase taxes to reduce consumption. Expert knowledge is required to determine if this is a problem.

5. Let $D2_i = 1$ if $i = 2$ and 0 otherwise; $D3_i = 1$ if $i = 3$ and 0 otherwise … $Dn_i = 1$ if $i = n$ and 0 otherwise. Let $B2_t = 1$ if $t = 2$ and 0 otherwise; $B3_t = 1$ if $t = 3$ and 0 otherwise … $BT_t = 1$ if $t = T$ and 0 otherwise. Let $\beta_0 = \alpha_1 + \mu_1$; $\gamma_i = \alpha_i - \alpha_1$ and $\delta_t = \mu_t - \mu_1$.

7. (a) Average snow fall does not vary over time, and thus will be perfectly collinear with the state fixed effect.
   (b) $Snow_{it}$ does vary with time, and so this method can be used along with state fixed effects.

9. This assumption is necessary for the usual formula for SEs to be correct. If it is incorrect, errors are correlated, the usual formula for SEs are wrong and inference is faulty. The appendix includes a discussion of more general formulae for the SEs when Assumption #5 is violated.

11. No, one of the regressors is Y_{it-1}. This depends on Y_{it-1}. This means that assumption (1) is violated.

Chapter 11
Regression with a Binary Dependent Variable

■ Solutions to Exercises

1. (a) The *t*-statistic for the coefficient on *Experience* is 0.031/0.009 = 3.44, which is significant at the 1% level.

   (b) $z_{Matthew} = 0.712 + 0.031 \times 10 = 1.022; \Phi(1.022) = 0.847$

   (c) $z_{Christopher} = 0.712 + 0.031 \times 0 = 0.712; \Phi(0.712) = 0.762$

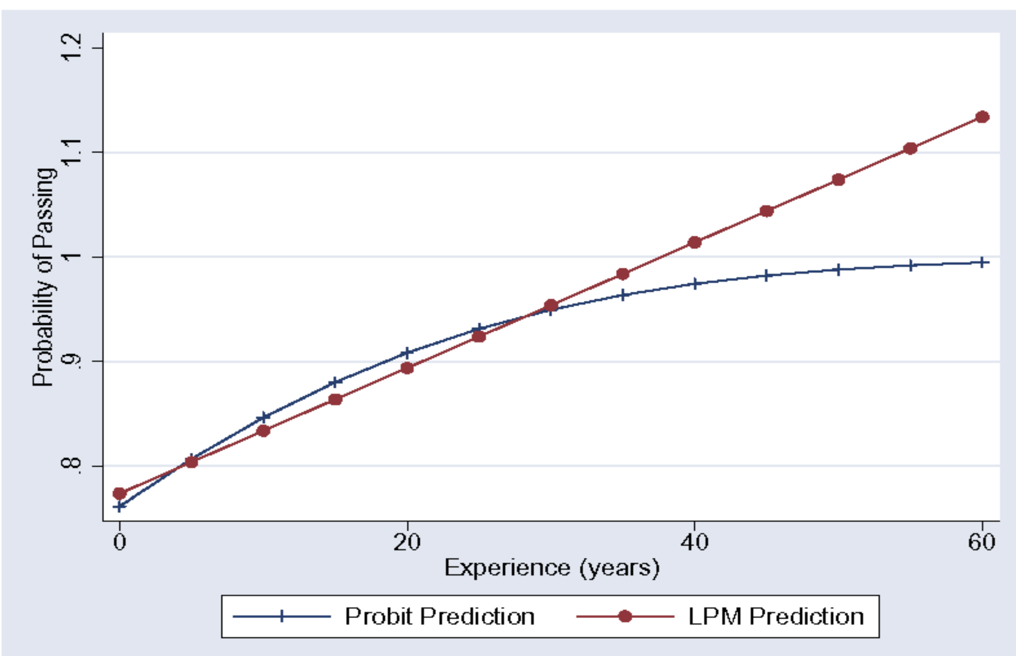
   (d) $z_{Jed} = 0.712 + 0.031 \times 80 = 3.192; \Phi(3.192) = 0.999,$ this is unlikely to be accurate because the sample did not include anyone with more that 40 years of driving experience.

3. (a) The *t*-statistic for the coefficient on *Experience* is $t = 0.006/0.002 = 3$, which is significant a the 1% level.

   (b) $Prob_{Matther} = 0.774 + 0.006 \times 10 = 0.836$

   (c) $Prob_{Christopher} = 0.774 + 0.006 \times 0 = 0.774$

   (d)



The probabilities are similar except when experience in large (>40 years). In this case the LPM model produces nonsensical results (probabilities greater than 1.0).

5. (a) $\Phi(0.806 + 0.041 \times 10 \times 0.174 \times 1 - 0.015 \times 1 \times 10) = 0.814$
   (b) $\Phi(0.806 + 0.041 \times 2 - 0.174 \times 0 - 0.015 \times 0 \times 2) = 0.813$
   (c) The *t*-stat on the interaction term is –0.015/0.019 = –0.79, which is insignificant at the 10% level.

7. (a) For a black applicant having a P/I ratio of 0.35, the probability that the application will be denied is $F(-4.13 + 5.37 \times 0.35 + 1.27) = \frac{1}{1+e^{0.9805}} = 27.28\%$.
   (b) With the P/I ratio reduced to 0.30, the probability of being denied is $F(-4.13 + 5.37 \times 0.30 + 1.27) = \frac{1}{1+e^{1.249}} = 22.29\%$. The difference in denial probabilities compared to (a) is 4.99 percentage points lower.
   (c) For a white applicant having a P/I ratio of 0.35, the probability that the application will be denied is $F(-4.13 + 5.37 \times 0.35) = \frac{1}{1+e^{2.2505}} = 9.53\%$. If the P/I ratio is reduced to 0.30, the probability of being denied is $F(-4.13 + 5.37 \times 0.30) = \frac{1}{1+e^{2.519}} = 7.45\%$. The difference in denial probabilities is 2.08 percentage points lower.
   (d) From the results in parts (a)–(c), we can see that the marginal effect of the P/I ratio on the probability of mortgage denial depends on race. In the logit regression functional form, the marginal effect depends on the level of probability which in turn depends on the race of the applicant. The coefficient on *black* is statistically significant at the 1% level. The logit and probit results are similar.

9. (a) The coefficient on *black* is 0.084, indicating an estimated denial probability that is 8.4 percentage points higher for the black applicant.
   (b) The 95% confidence interval is $0.084 \pm 1.96 \times 0.023 = [3.89\%, 12.91\%]$.
   (c) The answer in (a) will be biased if there are omitted variables which are race-related and have impacts on mortgage denial. Such variables would have to be related with race and also be related with the probability of default on the mortgage (which in turn would lead to denial of the mortgage application). Standard measures of default probability (past credit history and employment variables) are included in the regressions shown in Table 9.2, so these omitted variables are unlikely to bias the answer in (a). Other variables such as education, marital status, and occupation may also be related the probability of default, and these variables are omitted from the regression in column. Adding these variables (see columns (4)–(6)) have little effect on the estimated effect of *black* on the probability of mortgage denial.

11. (a) This is a censored or truncated regression model (note the dependent variable might be zero).
   (b) This is an ordered response model.
   (c) This is the discrete choice (or multiple choice) model.
   (d) This is a model with count data.

Chapter 12
Instrumental Variables Regression

■ Solutions to Exercises

1. (a) The change in the regressor, $\ln(P_{i,1995}^{cigarettes}) - \ln(P_{i,1985}^{cigarettes})$, from a \$0.10 per pack increase in the retail price is ln 2.10 − ln 2.00 = 0.0488. The expected percentage change in cigarette demand is −9.94 × 0.0488 × 100% = −4.5872%. The 95% confidence interval is (–0.94 ± 1.96 × 0.21) × 0.0488 × 100% = [–6.60%, –2.58%].
   (b) With a 2% reduction in income, the expected percentage change in cigarette demand is 0.53 × (−0.02) × 100% = −1.06%.
   (c) The regression in column (1) will not provide a reliable answer to the question in (b) when recessions last less than 1 year. The regression in column (1) studies the long-run price and income elasticity. Cigarettes are addictive. The response of demand to an income decrease will be smaller in the short run than in the long run.
   (d) The instrumental variable would be too weak (irrelevant) if the *F*-statistic in column (1) was 3.6 instead of 33.6, and we cannot rely on the standard methods for statistical inference. Thus the regression would not provide a reliable answer to the question posed in (a).

3. (a) The estimator $\hat{\sigma}_a^2 = \frac{1}{n-2}\sum_{i=1}^n (Y_i - \hat{\beta}_0^{TSLS} - \hat{\beta}_1^{TSLS}\hat{X}_i)^2$ is not consistent. Write this as $\hat{\sigma}_a^2 = \frac{1}{n-2}\sum_{i=1}^n (\hat{u}_i - \hat{\beta}_1^{TSLS}(\hat{X}_i - X_i))^2$, where $\hat{u}_i = Y_i - \hat{\beta}_0^{TSLS} - \hat{\beta}_1^{TSLS}X_i$. Replacing $\hat{\beta}_1^{TSLS}$ with β_1, as suggested in the question, write this as $\hat{\sigma}_a^2 \approx \frac{1}{n}\sum_{i=1}^n (u_i - \beta_1(\hat{X}_i - X_i))^2 = \frac{1}{n}\sum_{i=1}^n u_i^2 + \frac{1}{n}\sum_{i=1}^n [\beta_1^2(\hat{X}_i - X_i)^2 + 2u_i\beta_1(\hat{X}_i - X_i)]$. The first term on the right hand side of the equation converges to $\hat{\sigma}_u^2$, but the second term converges to something that is non-zero. Thus $\hat{\sigma}_a^2$ is not consistent.
   (b) The estimator $\hat{\sigma}_b^2 = \frac{1}{n-2}\sum_{i=1}^n (Y_i - \hat{\beta}_0^{TSLS} - \hat{\beta}_1^{TSLS}X_i)^2$ is consistent. Using the same notation as in (a), we can write $\hat{\sigma}_b^2 \approx \frac{1}{n}\Sigma_{i=1}^n u_i^2$, and this estimator converges in probability to σ_u^2.

5. (a) Instrument relevance. Z_i does not enter the population regression for X_i
   (b) Z is not a valid instrument. $\hat{X}^*$ will be perfectly collinear with W. (Alternatively, the first stage regression suffers from perfect multicollinearity.)
   (c) W is perfectly collinear with the constant term.
   (d) Z is not a valid instrument because it is correlated with the error term.

7. (a) Under the null hypothesis of instrument exogeneity, the J statistic is distributed as a χ_1^2 random variable, with a 1% critical value of 6.63. Thus the statistic is significant, and instrument exogeneity $E(u_i|Z_{1i}, Z_{2i}) = 0$ is rejected.
   (b) The J test suggests that $E(u_i|Z_{1i}, Z_{2i}) \neq 0$, but doesn't provide evidence about whether the problem is with Z_1 or Z_2 or both.

9. (a) There are other factors that could affect both the choice to serve in the military and annual earnings. One example could be education, although this could be included in the regression as a control variable. Another variable is “ability” which is difficult to measure, and thus difficult to control for in the regression.

   (b) The draft was determined by a national lottery so the choice of serving in the military was random. Because it was randomly selected, the lottery number is uncorrelated with individual characteristics that may affect earning and hence the instrument is exogenous. Because it affected the probability of serving in the military, the lottery number is relevant.

Chapter 13
Experiments and Quasi-Experiments

■ Solutions to Exercises

1. For students in kindergarten, the estimated small class treatment effect relative to being in a regular class is an increase of 13.90 points on the test with a standard error 2.45. The 95% confidence interval is 13.90 ± 1.96 × 2.45 = [9.098, 18.702].

   For students in grade 1, the estimated small class treatment effect relative to being in a regular class is an increase of 29.78 points on the test with a standard error 2.83. The 95% confidence interval is 29.78 ± 1.96 × 2.83 = [24.233, 35.327].

   For students in grade 2, the estimated small class treatment effect relative to being in a regular class is an increase of 19.39 points on the test with a standard error 2.71. The 95% confidence interval is 19.39 ± 1.96 × 2.71 = [14.078, 24.702].

   For students in grade 3, the estimated small class treatment effect relative to being in a regular class is an increase of 15.59 points on the test with a standard error 2.40. The 95% confidence interval is 15.59 ± 1.96 × 2.40 = [10.886, 20.294].

3. (a) The estimated average treatment effect is $\overline{X}_{TreatmentGroup} - \overline{X}_{Control}$ = 1241 – 1201 = 40 points.

   (b) There would be nonrandom assignment if men (or women) had different probabilities of being assigned to the treatment and control groups. Let p_{Men} denote the probability that a male is assigned to the treatment group. Random assignment means p_{Men} = 0.5. Testing this null hypothesis results in a t-statistic of $t_{Men} = \frac{\hat{p}_{Men} - 0.5}{\sqrt{\frac{1}{n_{men}}\hat{p}_{Men}(1-\hat{p}_{Men})}} = \frac{0.55-0.5}{\sqrt{\frac{1}{100}0.55(1-45)}} = 1.00,$ so that the null of random assignment cannot be rejected at the 10% level. A similar result is found for women.

5. (a) This is an example of attrition, which poses a threat to internal validity. After the male athletes leave the experiment, the remaining subjects are representative of a population that excludes male athletes. If the average causal effect for this population is the same as the average causal effect for the population that includes the male athletes, then the attrition does not affect the internal validity of the experiment. On the other hand, if the average causal effect for male athletes differs from the rest of population, internal validity has been compromised.

   (b) This is an example of partial compliance which is a threat to internal validity. The local area network is a failure to follow treatment protocol, and this leads to bias in the OLS estimator of the average causal effect.

   (c) This poses no threat to internal validity. As stated, the study is focused on the effect of dorm room Internet connections. The treatment is making the connections available in the room; the treatment is not the use of the Internet. Thus, the art majors received the treatment (although they chose not to use the Internet).

(d) As in part (b) this is an example of partial compliance. Failure to follow treatment protocol leads to bias in the OLS estimator.

7. From the population regression

$$Y_{it} = \alpha_i + \beta_1 X_{it} + \beta_2 (D_t \times W_i) + \beta_0 D_t + v_{it},$$

we have

$$Y_{i2} - Y_{i1} = \beta_1 (X_{i2} - X_{i1}) + \beta_2 [(D_2 - D_1) \times W_i] + \beta_0 (D_2 - D_1) + (v_{i2} - v_{i1}).$$

By defining $\Delta Y_i = Y_{i2} - Y_{i1}$, $\Delta X_i = X_{i2} - X_{i1}$ (a binary treatment variable) and $u_i = v_{i2} - v_{i1}$, and using $D_1 = 0$ and $D_2 = 1$, we can rewrite this equation as

$$\Delta Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + u_i,$$

which is Equation (13.5) in the case of a single W regressor.

9. The covariance between $\beta_{1i} X_i$ and X_i is

$$\begin{aligned} \text{cov}(\beta_{1i} X_i, X_i) &= E\{[\beta_{1i} X_i - E(\beta_{1i} X_i)][X_i - E(X_i)]\} \\ &= E\{\beta_{1i} X_i^2 - E(\beta_{1i} X_i) X_i - \beta_{1i} X_i E(X_i) + E(\beta_{1i} X_i) E(X_i)\} \\ &= E(\beta_{1i} X_i^2) - E(\beta_{1i} X_i) E(X_i) \end{aligned}$$

Because X_i is randomly assigned, X_i is distributed independently of β_{1i}. The independence means

$$E(\beta_{1i} X_i) = E(\beta_{1i}) E(X_i) \quad \text{and} \quad E(\beta_{1i} X_i^2) = E(\beta_{1i}) E(X_i^2).$$

Thus $\text{cov}(\beta_{1i} X_i, X_i)$ can be further simplified:

$$\begin{aligned} \text{cov}(\beta_{1i} X_i, X_i) &= E(\beta_{1i})[E(X_i^2) - E^2(X_i)] \\ &= E(\beta_{1i}) \sigma_X^2. \end{aligned}$$

So

$$\frac{\text{cov}(\beta_{1i} X_i, X_i)}{\sigma_X^2} = \frac{E(\beta_{1i}) \sigma_X^2}{\sigma_X^2} = E(\beta_{1i}).$$

11. Following the notation used in Chapter 13, let π_{1i} denote the coefficient on state sales tax in the "first stage" IV regression, and let $-\beta_{1i}$ denote cigarette demand elasticity. (In both cases, suppose that income has been controlled for in the analysis.) From (13.11)

$$\hat{\beta}^{TSLS} \xrightarrow{p} \frac{E(\beta_{1i}\ \pi_{1i})}{E(\pi_{1i})} = E(\beta_{1i}) + \frac{Cov(\beta_{1i}, \pi_{1i})}{E(\pi_{1i})} = \text{Average Treatment Effect} + \frac{Cov(\beta_{1i}, \pi_{1i})}{E(\pi_{1i})},$$

where the first equality uses the uses properties of covariances (equation (2.34)), and the second equality uses the definition of the average treatment effect. Evidently, the local average treatment effect will deviate from the average treatment effect when $Cov(\beta_{1i}, \pi_{1i}) \neq 0$. As discussed in Section 13.7, this covariance is zero when β_{1i} or π_{1i} are constant. This seems likely. But, for the sake of argument, suppose that they are not constant; that is, suppose the demand elasticity differs from state to state (β_{1i} is not constant) as does the effect of sales taxes on cigarette prices (π_{1i} is not constant). Are β_{1i} and π_{1i} related? Microeconomics suggests that might be. Recall from your microeconomics class that the lower is the demand elasticity, the larger fraction of a sales tax is passed along to consumers in terms of higher prices. This suggests that β_{1i} and π_{1i} are positively related, so that $Cov(\beta_{1i}, \pi_{1i}) > 0$. Because $E(\pi_{1i}) > 0$, this suggests that the local average treatment effect is greater than the average treatment effect when β_{1i} varies from state to state.

Chapter 14
Introduction to Time Series Regression and Forecasting

■ Solutions to Exercises

1. (a) Since the probability distribution of Y_t is the same as the probability distribution of Y_{t-1} (this is the definition of stationarity), the means (and all other moments) are the same.

   (b) $E(Y_t) = \beta_0 + \beta_1 E(Y_{t-1}) + E(u_t)$, but $E(u_t) = 0$ and $E(Y_t) = E(Y_{t-1})$. Thus $E(Y_t) = \beta_0 + \beta_1 E(Y_t)$, and solving for $E(Y_t)$ yields the result.

3. (a) To test for a stochastic trend (unit root) in ln(IP), the ADF statistic is the t-statistic testing the hypothesis that the coefficient on ln(IP_{t-1}) is zero versus the alternative hypothesis that the coefficient on ln(IP_{t-1}) is less than zero. The calculated t-statistic is $t = \frac{-0.018}{0.007} = -2.5714$. From Table 14.4, the 10% critical value with a time trend is –3.12. Because –2.5714 > –3.12, the test does not reject the null hypothesis that ln(IP) has a unit autoregressive root at the 10% significance level. That is, the test does not reject the null hypothesis that ln(IP) contains a stochastic trend, against the alternative that it is stationary.

   (b) The ADF test supports the specification used in Exercise 14.2. The use of first differences in Exercise 14.2 eliminates random walk trend in ln(IP).

5. (a)

$$\begin{aligned} E[(W-c)^2] &= E\{[W-\mu_W)+(\mu_W-c)]^2\} \\ &= E[(W-\mu_W)^2]+2E(W-\mu_W)(\mu_W-c)+(\mu_W-c)^2 \\ &= \sigma_W^2+(\mu_W-c)^2. \end{aligned}$$

   (b) Using the result in part (a), the conditional mean squared error

$$E[(Y_t - f_{t-1})^2 \mid Y_{t-1}, Y_{t-2}, \ldots] = \sigma^2_{t|t-1} + (Y_{t|t-1} - f_{t-1})^2$$

   with the conditional variance $\sigma^2_{t|t-1} = E[(Y_t - Y_{t|t-1})^2]$. This equation is minimized when the second term equals zero, or when $f_{t-1} = Y_{t|t-1}$.

   (c) Applying Equation (2.27), we know the error u_t is uncorrelated with u_{t-1} if $E(u_t | u_{t-1}) = 0$. From Equation (14.14) for the AR(p) process, we have

$$u_{t-1} = Y_{t-1} - \beta_0 - \beta_1 Y_{t-2} - \beta_2 Y_{t-3} - \cdots - \beta_p Y_{t-p-1} = f(Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p-1}),$$

   a function of Y_{t-1} and its lagged values. The assumption $E(u_t | Y_{t-1}, Y_{t-2}, \ldots) = 0$ means that conditional on Y_{t-1} and its lagged values, or any functions of Y_{t-1} and its lagged values, u_t has mean zero. That is,

$$E(u_t | u_{t-1}) = E[u_t \mid f(Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p-2})] = 0.$$

Thus u_t and u_{t-1} are uncorrelated. A similar argument shows that u_t and u_{t-j} are uncorrelated for all $j \geq 1$. Thus u_t is serially uncorrelated.

7. (a) From Exercise (14.1) $E(Y_t) = 2.5 + 0.7E(Y_{t-1}) + E(u_t)$, but $E(Y_t) = E(Y_{t-1})$ (stationarity) and $E(u_t) = 0$, so that $E(Y_t) = 2.5/(1–0.7)$. Also, because $Y_t = 2.5 + 0.7Y_{t-1} + u_t$, $var(Y_t) = 0.7^2 var(Y_{t-1}) + var(u_t) + 2 \times 0.7 \times cov(Y_{t-1}, u_t)$. But $cov(Y_{t-1}, u_t) = 0$ and $var(Y_t) = var(Y_{t-1})$ (stationarity), so that $var(Y_t) =$
$9/(1 - 0.7^2) = 17.647$.

(b) The 1st autocovariance is

$$\begin{aligned}\text{cov}(Y_t, Y_{t-1}) &= \text{cov}(2.5 + 0.7Y_{t-1} + u_t, Y_{t-1})\\ &= 0.7\,\text{var}(Y_{t-1}) + \text{cov}(u_t, Y_{t-1})\\ &= 0.7\sigma_Y^2\\ &= 0.7 \times 17.647 = 12.353.\end{aligned}$$

The 2nd autocovariance is

$$\begin{aligned}\text{cov}(Y_t, Y_{t-2}) &= \text{cov}[(1+0.7)2.5 + 0.7^2 Y_{t-2} + u_t + 0.7u_{t-1}, Y_{t-2}]\\ &= 0.7^2\,\text{var}(Y_{t-2}) + \text{cov}(u_t + 0.7u_{t-1}, Y_{t-2})\\ &= 0.7^2\sigma_Y^2\\ &= 0.7^2 \times 17.647 = 8.6471.\end{aligned}$$

(c) The 1st autocorrelation is

$$\text{corr}\,(Y_t, Y_{t-1}) = \frac{\text{cov}(Y_t, Y_{t-1})}{\sqrt{\text{var}(Y_t)\,\text{var}(Y_{t-1})}} = \frac{0.7\sigma_Y^2}{\sigma_Y^2} = 0.7.$$

The 2nd autocorrelation is

$$\text{corr}\,(Y_t, Y_{t-2}) = \frac{\text{cov}(Y_t, Y_{t-2})}{\sqrt{\text{var}(Y_t)\,\text{var}(Y_{t-2})}} = \frac{0.7^2\sigma_Y^2}{\sigma_Y^2} = 0.49.$$

(d) The conditional expectation of Y_{T+1} given Y_T is

$$Y_{T+1/T} = 2.5 + 0.7Y_T = 2.5 + 0.7 \times 102.3 = 74.11.$$

9. (a) $E(Y_t) = \beta_0 + E(e_t) + b_1E(e_{t-1}) + \cdots + b_qE(e_{t-q}) = \beta_0$ [because $E(e_t) = 0$ for all values of t].

(b)

$$\begin{aligned}\text{var}(Y_t) &= \text{var}(e_t) + b_1^2\,\text{var}(e_{t-1}) + \cdots + b_q^2\,\text{var}(e_{t-q}) + 2b_1\,\text{cov}(e_t, e_{t-1}) + \cdots + 2b_{q-1}b_q\,\text{cov}(e_{t-q+1}, e_{t-q})\\ &= \sigma_e^2(1 + b_1^2 + \cdots + b_q^2)\end{aligned}$$

because $\text{var}(e_t) = \sigma_e^2$ for all t and $\text{cov}(e_t, e_i) = 0$ for $i \neq t$.

(c) $Y_t = \beta_0 + e_t + b_1e_{t-1} + b_2e_{t-2} + \cdots + b_qe_{t-q}$ and $Y_{t-j} = \beta_0 + e_{t-j} + b_1e_{t-1-j} + b_2e_{t-2-j} + \cdots + b_qe_{t-q-j}$ and $\text{cov}(Y_t, Y_{t-j}) = \sum_{k=0}^{q}\sum_{m=0}^{q} b_k b_m \,\text{cov}(e_{t-k}, e_{t-j-m})$, where $b_0 = 1$. Notice that $\text{cov}(e_{t-k}, e_{t-j-m}) = 0$ for all terms in the sum.

(d) $\text{var}(Y_t) = \sigma_e^2\left(1 + b_1^2\right)$, $\text{cov}(Y_t, Y_{t-j}) = \sigma_e^2 b_1$, and $\text{cov}(Y_t, Y_{t-j}) = 0$ for $j > 1$.

11. Write the model as $Y_t - Y_{t-1} = \beta_0 + \beta_1(Y_{t-1} - Y_{t-2}) + u_t$. Rearranging yields $Y_t = \beta_0 + (1+\beta_1)Y_{t-1} - \beta_1 Y_{t-2} + u_t$.

Chapter 15
Estimation of Dynamic Causal Effects

■ Solutions to Exercises

1. (a) See the table below. β_i is the dynamic multiplier. With the 25% oil price jump, the predicted effect on output growth for the ith quarter is $25\beta_i$ percentage points.

Period ahead (i)	**Dynamic multiplier (β_i)**	**Predicted effect on output growth ($25\beta_i$)**	**95% confidence interval $25 \times [\beta_i \pm 1.96SE(\beta_i)]$**
0	−0.055	−1.375	[–4.021, 1.271]
1	−0.026	−0.65	[–3.443, 2.143]
2	−0.031	−0.775	[–3.127, 1.577]
3	−0.109	−2.725	[–4.783, –0.667]
4	−0.128	−3.2	[–5.797, –0.603]
5	0.008	0.2	[–1.025, 1.425]
6	0.025	0.625	[–1.727, 2.977]
7	−0.019	−0.475	[–2.386, 1.436]
8	0.067	1.675	[–0.015, 0.149]

(b) The 95% confidence interval for the predicted effect on output growth for the i'th quarter from the 25% oil price jump is $25 \times [\beta_i \pm 1.96SE(\beta_i)]$ percentage points. The confidence interval is reported in the table in (a).

(c) The predicted cumulative change in GDP growth over eight quarters is

$$25 \times (-0.055 - 0.026 - 0.031 - 0.109 - 0.128 + 0.008 + 0.025 - 0.019) = -8.375\%.$$

(d) The 1% critical value for the F-test is 2.407. Since the HAC F-statistic 3.49 is larger than the critical value, we reject the null hypothesis that all the coefficients are zero at the 1% level.

3. The dynamic causal effects are for experiment A. The regression in exercise 15.1 does not control for interest rates, so that interest rates are assumed to evolve in their "normal pattern" given changes in oil prices.

5. Substituting

$$\begin{aligned}X_t &= \Delta X_t + X_{t-1} = \Delta X_t + \Delta X_{t-1} + X_{t-2} \\ &= \cdots\cdots \\ &= \Delta X_t + \Delta X_{t-1} + \cdots + \Delta X_{t-p+1} + X_{t-p}\end{aligned}$$

into Equation (15.4), we have

$$\begin{aligned}Y_t &= \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \cdots + \beta_{r+1} X_{t-r} + u_t \\ &= \beta_0 + \beta_1(\Delta X_t + \Delta X_{t-1} + \cdots + \Delta X_{t-r+1} + X_{t-r}) \\ &\quad + \beta_2(\Delta X_{t-1} + \cdots + \Delta X_{t-r+1} + X_{t-r}) \\ &\quad + \cdots + \beta_r(\Delta X_{t-r+1} + X_{t-r}) + \beta_{r+1} X_{t-r} + u_t \\ &= \beta_0 + \beta_1 \Delta X_t + (\beta_1 + \beta_2)\Delta X_{t-1} + (\beta_1 + \beta_2 + \beta_3)\Delta X_{t-2} \\ &\quad + \cdots + (\beta_1 + \beta_2 + \cdots + \beta_r)\Delta X_{t-r+1} \\ &\quad + (\beta_1 + \beta_2 + \cdots + \beta_r + \beta_{r+1}) X_{t-r} + u_t.\end{aligned}$$

Comparing the above equation to Equation (15.7), we see $\delta_0 = \beta_0$, $\delta_1 = \beta_1$, $\delta_2 = \beta_1 + \beta_2$, $\delta_3 = \beta_1 + \beta_2 + \beta_3, \ldots$, and $\delta_{r+1} = \beta_1 + \beta_2 + \cdots + \beta_r + \beta_{r+1}$.

7. Write $u_t = \sum_{i=0}^{\infty} \phi_1^i \tilde{u}_{t-i}$

   (a) Because $E(\tilde{u}_i | X_t) = 0$ for all i and t, $E(u_i | X_t) = 0$ for all i and t, so that X_t is strictly exogenous.

   (b) Because $E(u_{t-j} | \tilde{u}_{t+1}) = 0$ for $j \geq 0$, X_t is exogenous. However $E(u_{t+1} | \tilde{u}_{t+1}) = \tilde{u}_{t+1}$ so that X_t is not strictly exogenous.

9. (a) This follows from the material around equation (3.2).

   (b) Quasi differencing the equation yields $Y_t - \phi_1 Y_{t-1} = (1 - \phi_1)\beta_0 + \tilde{u}_t$, and the GLS estimator of $(1 - \phi_1)\beta_0$ is the mean of $Y_t - \phi_1 Y_{t-1} = \frac{1}{T-1}\sum_{t=2}^{T}(Y_t - \phi_1 Y_{t-1})$. Dividing by $(1 - \phi_1)$ yields the GLS estimator of β_0.

   (c) This is a rearrangement of the result in (b).

   (d) Write $\hat{\beta}_0 = \frac{1}{T}\sum_{t=1}^{T} Y_t = \frac{1}{T}(Y_T + Y_1) + \frac{T-1}{T}\frac{1}{T-1}\sum_{t=2}^{T-1} Y_t$, so that $\hat{\beta}_0 - \hat{\beta}_0^{GLS} = \frac{1}{T}(Y_T + Y_1) - \frac{1}{T}\frac{1}{T-1}\sum_{t=2}^{T-1} Y_t - \frac{1}{1-\phi}\frac{1}{T-1}(Y_T - Y_1)$ and the variance is seen to be proportional to $\frac{1}{T^2}$.

Chapter 16
Additional Topics in Time Series Regression

Solutions to Exercises

1. Y_t follows a stationary AR(1) model, $Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$. The mean of Y_t is $\mu_Y = E(Y_t) = \frac{\beta_0}{1-\beta_1}$, and $E(u_t|Y_t) = 0$.

   (a) The h-period ahead forecast of $Y_t, Y_{t+h|t} = E(Y_{t+h}|Y_t, Y_{t-1}, \ldots)$, is

$$
\begin{aligned}
Y_{t+h|t} = E(Y_{t+h}|Y_t, Y_{t-1}, \ldots) &= E(\beta_0 + \beta_1 Y_{t+h-1} + u_t|Y_t, Y_{t-1}, \ldots) \\
&= \beta_0 + \beta_1 Y_{t+h-1|t} = \beta_0 + \beta_1(\beta_0 + \beta_1 Y_{t+h-2|t}) \\
&= (1+\beta_1)\beta_0 + \beta_1^2 Y_{t+h-2|t} \\
&= (1+\beta_1)\beta_0 + \beta_1^2(\beta_0 + \beta_1 Y_{t+h-3|t}) \\
&= (1+\beta_1+\beta_1^2)\beta_0 + \beta_1^3 Y_{t+h-3|t} \\
&= \cdots\cdots \\
&= \left(1+\beta_1+\cdots+\beta_1^{h-1}\right)\beta_0 + \beta_1^h Y_t \\
&= \frac{1-\beta_1^h}{1-\beta_1}\beta_0 + \beta_1^h Y_t \\
&= \mu_Y + \beta_1^h (Y_t - \mu_Y).
\end{aligned}
$$

   (b) Substituting the result from part (a) into X_t gives

$$
\begin{aligned}
X_t &= \sum_{i=0}^{\infty} \delta^i Y_{t+i|t} = \sum_{i=0}^{\infty} \delta^i [\mu_Y + \beta_1^i (Y_t - \mu_Y)] \\
&= \mu_Y \sum_{i=0}^{\infty} \delta^i + (Y_t - \mu_Y) \sum_{i=0}^{\infty} (\beta_1 \delta)^i \\
&= \frac{\mu_Y}{1-\delta} + \frac{Y_t - \mu_Y}{1-\beta_1 \delta}.
\end{aligned}
$$

3. u_t follows the ARCH process with mean $E(u_t) = 0$ and variance $\sigma_t^2 = 1.0 + 0.5u_{t-1}^2$.

(a) For the specified ARCH process, u_t has the conditional mean $E(u_t|u_{t-1}) = 0$ and the conditional variance.

$$\text{var}(u_t|u_{t-1}) = \sigma_t^2 = 1.0 + 0.5u_{t-1}^2.$$

The unconditional mean of u_t is $E(u_t) = 0$, and the unconditional variance of u_t is

$$\begin{aligned}\text{var}(u_t) &= \text{var}[E(u_t|u_{t-1})] + E[\text{var}(u_t|u_{t-1})]\\ &= 0 + 1.0 + 0.5E(u_{t-1}^2)\\ &= 1.0 + 0.5\,\text{var}(u_{t-1}).\end{aligned}$$

The last equation has used the fact that $E(u_t^2) = \text{var}(u_t) + E(u_t)]^2 = \text{var}(u_t),$ which follows because $E(u_t) = 0$. Because of the stationarity, var(u_{t-1}) = var(u_t). Thus, var(u_t) = 1.0 + 0.5var(u_t) which implies $\text{var}(u_t) = \frac{1.0}{0.5} = 2.$

(b) When $u_{t-1} = 0.2,\ \sigma_t^2 = 1.0 + 0.5 \times 0.2^2 = 1.02.$ The standard deviation of u_t is $\sigma_t = 1.01$. Thus

$$\begin{aligned}\Pr(-3 \le u_t \le 3) &= \Pr\left(\frac{-3}{1.01} \le \frac{u_t}{\sigma_t} \le \frac{3}{1.01}\right)\\ &= \Phi(2.9703) - \Phi(-2.9703) = 0.9985 - 0.0015 = 0.9970.\end{aligned}$$

When $u_{t-1} = 2.0,\ \sigma_t^2 = 1.0 + 0.5 \times 2.0^2 = 3.0.$ The standard deviation of u_t is $\sigma_t = 1.732$. Thus

$$\begin{aligned}\Pr(-3 \le u_t \le 3) &= \Pr\left(\frac{-3}{1.732} \le \frac{u_t}{\sigma_t} \le \frac{3}{1.732}\right)\\ &= \Phi(1.732) - \Phi(-1.732) = 0.9584 - 0.0416 = 0.9168.\end{aligned}$$

5. Because $Y_t = Y_t - Y_{t-1} + Y_{t-1} = Y_{t-1} + \Delta Y_t,$

$$\sum_{t=1}^{T} Y_t^2 = \sum_{t=1}^{T} (Y_{t-1} + \Delta Y_t)^2 = \sum_{t=1}^{T} Y_{t-1}^2 + \sum_{t=1}^{T} (\Delta Y_t)^2 + 2\sum_{t=1}^{T} Y_{t-1}\Delta Y_t.$$

So

$$\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}\Delta Y_t = \frac{1}{T} \times \frac{1}{2}\left[\sum_{t=1}^{T} Y_t^2 - \sum_{t=1}^{T} Y_{t-1}^2 - \sum_{t=1}^{T} (\Delta Y_t)^2\right].$$

Note that $\sum_{t=1}^{T} Y_t^2 - \sum_{t=1}^{T} Y_{t-1}^2 = \left(\sum_{t=1}^{T-1} Y_t^2 + Y_T^2\right) - \left(Y_0^2 + \sum_{t=1}^{T-1} Y_t^2\right) = Y_T^2 - Y_0^2 = Y_T^2$ because $Y_0 = 0$. Thus:

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}\Delta Y_t &= \frac{1}{T} \times \frac{1}{2}\left[Y_T^2 - \sum_{t=1}^{T} (\Delta Y_t)^2\right]\\ &= \frac{1}{2}\left[\left(\frac{Y_T}{\sqrt{T}}\right)^2 - \frac{1}{T}\sum_{t=1}^{T} (\Delta Y_t)^2\right].\end{aligned}$$

7. $\hat{\beta} = \frac{\sum_{t=1}^{T} Y_t X_t}{\sum_{t=1}^{T} X_t^2} = \frac{\sum_{t=1}^{T} Y_t \Delta Y_{t+1}}{\sum_{t=1}^{T} (\Delta Y_{t+1})^2} = \frac{\frac{1}{T}\sum_{t=1}^{T} Y_t \Delta Y_{t+1}}{\frac{1}{T}\sum_{t=1}^{T} (\Delta Y_{t+1})^2}$. Following the hint, the numerator is the same expression as (16.21) (shifted forward in time 1 period), so that $\frac{1}{T}\sum_{t=1}^{T} Y_t \Delta Y_{t+1} \xrightarrow{d} \frac{\sigma_u^2}{2}(\chi_1^2 - 1)$. The denominator is $\frac{1}{T}\sum_{t=1}^{T} (\Delta Y_{t+1})^2 = \frac{1}{T}\sum_{t=1}^{T} u_{t+1}^2 \xrightarrow{p} \sigma_u^2$ by the law of large numbers. The result follows directly.

9. (a) From the law of iterated expectations

$$\begin{aligned} E(u_t^2) &= E\left(\sigma_t^2\right) \\ &= E\left(\alpha_0 + \alpha_1 u_{t-1}^2\right) \\ &= \alpha_0 + \alpha_1 E\left(u_{t-1}^2\right) \\ &= \alpha_0 + \alpha_1 E\left(u_t^2\right) \end{aligned}$$

where the last line uses stationarity of *u*. Solving for $E(u_t^2)$ gives the required result.

(b) As in (a)

$$\begin{aligned} E(u_t^2) &= E\left(\sigma_t^2\right) \\ &= E\left(\alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_p u_{t-p}^2\right) \\ &= \alpha_0 + \alpha_1 E\left(u_{t-1}^2\right) + \alpha_2 E\left(u_{t-2}^2\right) + \cdots + \alpha_p E\left(u_{t-p}^2\right) \\ &= \alpha_0 + \alpha_1 E\left(u_t^2\right) + \alpha_2 E\left(u_t^2\right) + \cdots + \alpha_p E\left(u_t^2\right) \end{aligned}$$

so that $E(u_t^2) = \dfrac{\alpha_0}{1 - \sum_{t=1}^{p} \alpha_i}$

(c) This follows from (b) and the restriction that $E(u_t^2) > 0$.

(d) As in (a)

$$\begin{aligned} E(u_t^2) &= E\left(\sigma_t^2\right) \\ &= \alpha_0 + \alpha_1 E\left(u_{t-1}^2\right) + \phi_1 E\left(\sigma_{t-1}^2\right) \\ &= \alpha_0 + (\alpha_1 + \phi_1) E\left(u_{t-1}^2\right) \\ &= \alpha_0 + (\alpha_1 + \phi_1) E\left(u_t^2\right) \\ &= \frac{\alpha_0}{1 - \alpha_1 - \phi_1} \end{aligned}$$

(e) This follows from (d) and the restriction that $E\left(u_t^2\right) > 0$.

Chapter 17
The Theory of Linear Regression with One Regressor

■ Solutions to Exercises

1. (a) Suppose there are n observations. Let b_1 be an arbitrary estimator of β_1. Given the estimator b_1, the sum of squared errors for the given regression model is

$$\sum_{i=1}^{n}(Y_i - b_1 X_i)^2.$$

$\hat{\beta}_1^{RLS}$, the restricted least squares estimator of β_1, minimizes the sum of squared errors. That is, $\hat{\beta}_1^{RLS}$ satisfies the first order condition for the minimization which requires the differential of the sum of squared errors with respect to b_1 equals zero:

$$\sum_{i=1}^{n} 2(Y_i - b_1 X_i)(-X_i) = 0.$$

Solving for b_1 from the first order condition leads to the restricted least squares estimator

$$\hat{\beta}_1^{RLS} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}.$$

(b) We show first that $\hat{\beta}_1^{RLS}$ is unbiased. We can represent the restricted least squares estimator $\hat{\beta}_1^{RLS}$ in terms of the regressors and errors:

$$\hat{\beta}_1^{RLS} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2} = \frac{\sum_{i=1}^{n} X_i(\beta_1 X_i + u_i)}{\sum_{i=1}^{n} X_i^2} = \beta_1 + \frac{\sum_{i=1}^{n} X_i u_i}{\sum_{i=1}^{n} X_i^2}.$$

Thus

$$E(\hat{\beta}_1^{RLS}) = \beta_1 + E\left(\frac{\sum_{i=1}^{n} X_i u_i}{\sum_{i=1}^{n} X_i^2}\right) = \beta_1 + E\left[\frac{\sum_{i=1}^{n} X_i E(u_i|X_1,\ldots,X_n)}{\sum_{i=1}^{n} X_i^2}\right] = \beta_1,$$

where the second equality follows by using the law of iterated expectations, and the third equality follows from

$$\frac{\sum_{i=1}^{n} X_i E(u_i | X_1, \ldots, X_n)}{\sum_{i=1}^{n} X_i^2} = 0$$

because the observations are i.i.d. and $E(u_i | X_i) = 0$. (Note, $E(u_i | X_1, \ldots, X_n) = E(u_i | X_i)$ because the observations are i.i.d.

Under assumptions 1–3 of Key Concept 17.1, $\hat{\beta}_1^{RLS}$ is asymptotically normally distributed. The large sample normal approximation to the limiting distribution of $\hat{\beta}_1^{RLS}$ follows from considering

$$\hat{\beta}_1^{RLS} - \beta_1 = \frac{\sum_{i=1}^n X_i u_i}{\sum_{i=1}^n X_i^2} = \frac{\frac{1}{n}\sum_{i=1}^n X_i u_i}{\frac{1}{n}\sum_{i=1}^n X_i^2}.$$

Consider first the numerator which is the sample average of $v_i = X_i u_i$. By assumption 1 of Key Concept 17.1, v_i has mean zero: $E(X_i u_i) = E[X_i E(u_i|X_i)] = 0.$ By assumption 2, v_i is i.i.d. By assumption 3, var(v_i) is finite. Let $\bar{v} = \frac{1}{n}\sum_{i=1}^n X_i u_{i,}$ then $\sigma_{\bar{v}}^2 = \sigma_v^2/n.$ Using the central limit theorem, the sample average

$$\bar{v}/\sigma_{\bar{v}} = \frac{1}{\sigma_v \sqrt{n}} \sum_{i=1}^n v_i \xrightarrow{d} N(0, 1)$$

or

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i u_i \xrightarrow{d} N(0, \sigma_v^2).$$

For the denominator, X_i^2 is i.i.d. with finite second variance (because X has a finite fourth moment), so that by the law of large numbers

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{p} E(X^2).$$

Combining the results on the numerator and the denominator and applying Slutsky's theorem lead to

$$\sqrt{n}(\hat{\beta}_1^{RLS} - \beta_u) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i u_i}{\frac{1}{n}\sum_{i=1}^n X_i^2} \xrightarrow{d} N\left(0, \frac{\text{var}(X_i u_i)}{E(X^2)}\right).$$

(c) $\hat{\beta}_1^{RLS}$ is a linear estimator:

$$\hat{\beta}_1^{RLS} = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2} = \sum_{i=1}^n a_i Y_i, \qquad \text{where } a_i = \frac{X_i}{\sum_{i=1}^n X_i^2}.$$

The weight a_i ($i = 1,\ldots, n$) depends on $X_1,\ldots, X_n$ but not on $Y_1,\ldots, Y_n$.
Thus

$$\hat{\beta}_1^{RLS} = \beta_1 + \frac{\sum_{i=1}^n X_i u_i}{\sum_{i=1}^n X_i^2}.$$

$\hat{\beta}_1^{RLS}$ is conditionally unbiased because

$$
\begin{aligned}
E(\hat{\beta}_1^{RLS}|X_1,\ldots,X_n &= E\left(\beta_1 + \frac{\sum_{i=1}^n X_i u_i}{\sum_{i=1}^n X_i^2}|X_1,\ldots,X_n\right) \\
&= \beta_1 + E\left(\frac{\sum_{i=1}^n X_i u_i}{\sum_{i=1}^n X_i^2}|X_1,\ldots,X_n\right) \\
&= \beta_1.
\end{aligned}
$$

The final equality used the fact that

$$E\left(\frac{\sum_{i=1}^{n} X_i u_i}{\sum_{i=1}^{n} X_i^2}|X_1,\ldots, X_n\right) = \frac{\sum_{i=1}^{n} X_i E(u_i|X_1,\ldots, X_n)}{\sum_{i=1}^{n} X_i^2} = 0$$

because the observations are i.i.d. and E $(u_i|X_i) = 0$.

(d) The conditional variance of $\hat{\beta}_1^{RLS}$, given $X_1,\ldots, X_n$, is

$$\begin{aligned}\text{var}(\hat{\beta}_1^{RLS}|X1,\ldots, X_n) &= \text{var}\left(\beta_1 + \frac{\sum_{i=1}^{n} X_i u_i}{\sum_{i=1}^{n} X_i^2}|X_1,\ldots, X_n\right) \\ &= \frac{\sum_{i=1}^{n} X_i^2 \text{ var}(u_i|X_1,\ldots, X_n)}{(\sum_{i=1}^{n} X_i^2)^2} \\ &= \frac{\sum_{i=1}^{n} X_i^2 \sigma_u^2}{(\sum_{i=1}^{n} X_i^2)^2} \\ &= \frac{\sigma_u^2}{\sum_{i=1}^{n} X_i^2}.\end{aligned}$$

(e) The conditional variance of the OLS estimator $\hat{\beta}_1$ is

$$\text{var}(\hat{\beta}_1|X_1,\ldots, X_n) = \frac{\sigma_u^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Since

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - 2\bar{X}\sum_{i=1}^{n} X_i + n\bar{X}^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 < \sum_{i=1}^{n} X_i^2,$$

the OLS estimator has a larger conditional variance: $\text{var}(\bar{\beta}_1|X_1,\ldots, X_n) > \text{var}(\hat{\beta}_1^{RLS}|X_1,\ldots, X_n)$. The restricted least squares estimator $\hat{\beta}_1^{RLS}$ is more efficient.

(f) Under assumption 5 of Key Concept 17.1, conditional on $X_1,\ldots, X_n$, $\hat{\beta}_1^{RLS}$ is normally distributed since it is a weighted average of normally distributed variables u_i:

$$\hat{\beta}_1^{RLS} = \beta_1 + \frac{\sum_{i=1}^{n} X_i u_i}{\sum_{i=1}^{n} X_i^2}.$$

Using the conditional mean and conditional variance of $\hat{\beta}_1^{RLS}$ derived in parts (c) and (d) respectively, the sampling distribution of $\hat{\beta}_1^{RLS}$, conditional on $X_1,\ldots, X_n$, is

$$\hat{\beta}_1^{RLS} \sim N\left(\beta_1, \frac{\sigma_u^2}{\sum_{i=1}^{n} X_i^2}\right).$$

(g) The estimator

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} X_i} = \frac{\sum_{i=1}^{n}(\beta_1 X_i + u_i)}{\sum_{i=1}^{n} X_i} = \beta_1 + \frac{\sum_{i=1}^{n} u_i}{\sum_{i=1}^{n} X_i}$$

The conditional variance is

$$\begin{aligned}\operatorname{var}(\tilde{\beta}_1|X_1,\ldots,X_n) &= \operatorname{var}\left(\beta_1 + \frac{\sum_{i=1}^n u_i}{\sum_{i=1}^n X_i}|X_1,\ldots,X_n\right)\\ &= \frac{\sum_{i=1}^n \operatorname{var}(u_i|X_1,\ldots,X_n)}{(\sum_{i=1}^n X_i)^2}\\ &= \frac{n\sigma_u^2}{(\sum_{i=1}^n X_i)^2}.\end{aligned}$$

The difference in the conditional variance of $\tilde{\beta}_1$ and $\hat{\beta}_1^{RLS}$ is

$$\operatorname{var}(\tilde{\beta}_1|X_1,\ldots,X_n) - \operatorname{var}(\hat{\beta}_1^{RLS}|X_1,\ldots,X_n) = \frac{n\sigma_u^2}{(\sum_{i=1}^n X_i)^2} - \frac{\sigma_u^2}{\sum_{i=1}^n X_i^2}.$$

In order to prove $\operatorname{var}(\tilde{\beta}_1|X_1,\ldots,X_n) \geq \operatorname{var}(\hat{\beta}_1^{RLS}|X_1,\ldots,X_n),$ we need to show

$$\frac{n}{(\sum_{i=1}^n X_i)^2} \geq \frac{1}{\sum_{i=1}^n X_i^2}$$

or equivalently

$$n\sum_{i=1}^n X_i^2 \geq \left(\sum_{i=1}^n X_i\right)^2.$$

This inequality comes directly by applying the Cauchy-Schwartz inequality

$$\left[\sum_{i=1}^n (a_i \cdot b_i)\right]^2 \leq \sum_{i=1}^n a_i^2 \cdot \sum_{i=1}^n b_i^2$$

which implies

$$\left(\sum_{i=1}^n X_i\right)^2 = \left(\sum_{i=1}^n 1\cdot X_i\right)^2 \leq \sum_{i=1}^n 1^2 \cdot \sum_{i=1}^n X_i^2 = n\sum_{i=1}^n X_i^2.$$

That is $n\Sigma_{i=1}^n X_i^2 \geq (\Sigma_{x=1}^n X_i)^2,$ or $\operatorname{var}(\tilde{\beta}_1|X_1,\ldots,X_n) \geq \operatorname{var}(\hat{\beta}_1^{RLS}|X_1,\ldots,X_n).$

Note: because $\tilde{\beta}_1$ is linear and conditionally unbiased, the result $\operatorname{var}(\tilde{\beta}_1|X_1,\ldots,X_n) \geq \operatorname{var}(\hat{\beta}_1^{RLS}|X_1,\ldots,X_n)$ follows directly from the Gauss-Markov theorem.

3. (a) Using Equation (17.19), we have

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) = \sqrt{n}\,\frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})u_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$
$$= \sqrt{n}\,\frac{\frac{1}{n}\sum_{i=1}^{n}[(X_i - \mu_X) - (\overline{X} - \mu_X)]u_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$
$$= \frac{\sqrt{\frac{1}{n}}\sum_{i=1}^{n}(X_i - \mu_X)u_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2} - \frac{(\overline{X} - \mu_X)\sqrt{\frac{1}{n}}\sum_{i=1}^{n}u_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$
$$= \frac{\sqrt{\frac{1}{n}}\sum_{i=1}^{n}v_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2} - \frac{(\overline{X} - \mu_X)\sqrt{\frac{1}{n}\sum_{i=1}^{n}u_i}}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

by defining $v_i = (X_i - \mu_X)u_i$.

(b) The random variables $u_1, \ldots, u_n$ are i.i.d. with mean $\mu_u = 0$ and variance $0 < \sigma_u^2 < \infty$. By the central limit theorem,

$$\frac{\sqrt{n}(\overline{u} - \mu_u)}{\sigma_u} = \frac{\sqrt{\frac{1}{n}}\sum_{i=1}^{n}u_i}{\sigma_u} \xrightarrow{d} N(0, 1).$$

The law of large numbers implies $\overline{X} \xrightarrow{p} \mu_{X_2}$, or $\overline{X} - \mu_X \xrightarrow{p} 0$. By the consistency of sample variance, $\frac{1}{n}\Sigma_{i=1}^{n}(X_i - \overline{X})^2$ converges in probability to population variance, var(X_i), which is finite and non-zero. The result then follows from Slutsky's theorem.

(c) The random variable $v_i = (X_i - \mu_X)\, u_i$ has finite variance:

$$\text{var}(v_i) = \text{var}[(X_i - \mu_X)\mu_i]$$
$$\leq E[(X_i - \mu_X)^2 u_i^2]$$
$$\leq \sqrt{E[(X_i - \mu_X)^4]E[(u_i)^4]} < \infty.$$

The inequality follows by applying the Cauchy-Schwartz inequality, and the second inequality follows because of the finite fourth moments for (X_i, u_i). The finite variance along with the fact that v_i has mean zero (by assumption 1 of Key Concept 15.1) and v_i is i.i.d. (by assumption 2) implies that the sample average $\overline{v}$ satisfies the requirements of the central limit theorem. Thus,

$$\frac{\overline{v}}{\sigma_{\overline{v}}} = \frac{\sqrt{\frac{1}{n}}\sum_{i=1}^{n}v_i}{\sigma_v}$$

satisfies the central limit theorem.

(d) Applying the central limit theorem, we have

$$\frac{\sqrt{\frac{1}{n}}\sum_{i=1}^{n}v_i}{\sigma_v} \xrightarrow{d} N(0, 1).$$

Because the sample variance is a consistent estimator of the population variance, we have

$$\frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}{\text{var}(X_i)} \xrightarrow{p} 1.$$

Using Slutsky's theorem,

$$\frac{\dfrac{\frac{1}{n}\sum_{i=1}^{n} v_t}{\sigma_v}}{\dfrac{\frac{1}{n}\sum_{i=1}^{n}(X_t - \overline{X})^2}{\sigma_X^2}} \xrightarrow{d} N(0,1),$$

or equivalently

$$\frac{\sqrt{\frac{1}{n}}\sum_{i=1}^{n} v_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2} \xrightarrow{d} N\left(0,\ \frac{\operatorname{var}(v_i)}{[\operatorname{var}(X_i)]^2}\right).$$

Thus

$$\begin{aligned}\sqrt{n}(\hat{\beta}_1 - \beta_1) &= \frac{\sqrt{\frac{1}{n}}\sum_{i=1}^{n} v_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2} - \frac{(\overline{X} - \mu_X)\sqrt{\frac{1}{n}}\sum_{i=1}^{n} u_i}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2} \\ &\xrightarrow{d} N\left(0,\ \frac{\operatorname{var}(v_i)}{[\operatorname{var}(X_i)]^2}\right)\end{aligned}$$

since the second term for $\sqrt{n}(\hat{\beta}_1 - \beta_1)$ converges in probability to zero as shown in part (b).

5. Because $E(W^4) = [E(W^2)]^2 + \operatorname{var}(W^2)$, $[E(W^2)]^2 \le E(W^4) < \infty$. Thus $E(W^2) < \infty$.

7. (a) The joint probability distribution function of u_i, u_j, X_i, X_j is $f(u_i, u_j, X_i, X_j)$. The conditional probability distribution function of u_i and X_i given u_j and X_j is $f(u_i, X_i | u_j, X_j)$. Since u_i, X_i, $i = 1,\dots, n$ are i.i.d., $f(u_i, X_i | u_j, X_j) = f(u_i, X_i)$. By definition of the conditional probability distribution function, we have

$$\begin{aligned}f(u_i, u_j, X_i, X_j) &= f(u_i, X_i | u_j, X_j) f(u_j, X_j) \\ &= f(u_i, X_i) f(u_j, X_j).\end{aligned}$$

(b) The conditional probability distribution function of u_i and u_j given X_i and X_j equals

$$f(u_i, u_j | X_i, X_j) = \frac{f(u_i, u_j, X_i, X_j)}{f(X_i, X_j)} = \frac{f(u_i, X_i) f(u_j, X_j)}{f(X_i) f(X_j)} = f(u_i | X_i) f(u_j | X_j).$$

The first and third equalities used the definition of the conditional probability distribution function. The second equality used the conclusion the from part (a) and the independence between X_i and X_j. Substituting

$$f(u_i, u_j | X_i, X_j) = f(u_i | X_i) f(u_j | X_j)$$

into the definition of the conditional expectation, we have

$$\begin{aligned} E(u_i u_j|X_i, X_j) &= \int\int u_i u_j f(u_i, u_j|X_i, X_j)\, du_i du_j \\ &= \int\int u_i u_j f(u_i|X_i) f(u_j|X_j) du_i du_j \\ &= \int u_i f(u_i|X_i) du_i \int u_j f(u_j|X_j) du_j \\ &= E(u_i|X_i) E(u_j|X_j). \end{aligned}$$

(c) Let $Q = (X_1, X_2,\ldots, X_{i-1}, X_{i+1},\ldots, X_n)$, so that $f(u_i|X_1,\ldots, X_n) = f(u_i|X_i, Q)$. Write

$$\begin{aligned} f(u_i|X_i, Q) &= \frac{f(u_i, X_i, Q)}{f(X_i, Q)} \\ &= \frac{f(u_i, X_i) f(Q)}{f(X_i) f(Q)} \\ &= \frac{f(u_i, X_i)}{f(X_i)} \\ &= f(u_i|X_i) \end{aligned}$$

where the first equality uses the definition of the conditional density, the second uses the fact that (u_i, X_i) and Q are independent, and the final equality uses the definition of the conditional density. The result then follows directly.

(d) An argument like that used in (c) implies

$$f(u_i u_j|X_i,\ldots X_n) = f(u_i u_j|X_i, X_j)$$

and the result then follows from part (b).

9. We need to prove

$$\frac{1}{n}\sum_{i=1}^{n}[(X_i - \overline{X})^2 \hat{u}_i^2 - (X_i - \mu_X)^2 u_i^2] \xrightarrow{p} 0.$$

Using the identity $\overline{X} = \mu_X + (\overline{X} - \mu_X)$,

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n}[(X_i - \overline{X})^2 \hat{u}_i^2 - (X_i - \mu_X)^2 u_i^2] &= (\overline{X} - \mu_X)^2 \frac{1}{n}\sum_{i=1}^{n}\hat{u}_i^2 \\ &\quad - 2(\overline{X} - \mu_X)\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_X)\hat{u}_i^2 \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_X)^2(\hat{u}_i^2 - u_i^2). \end{aligned}$$

The definition of $\hat{u}_i$ implies

$$\begin{aligned} \hat{u}_i^2 &= u_i^2 + (\hat{\beta}_0 - \beta_0)^2 + (\hat{\beta}_1 - \beta_1)^2 X_i^2 - 2u_i(\hat{\beta}_0 - \beta_0) \\ &\quad - 2u_i(\hat{\beta}_1 - \beta_1)X_i + 2(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1)X_i. \end{aligned}$$

Substituting this into the expression for $\frac{1}{n}\Sigma_{i=1}^{n}[(X_i - \overline{X})^2 \hat{u}_i^2 - (X_i - \mu_X)^2 u_i^2]$ yields a series of terms each of which can be written as $a_n b_n$ where $a_n \xrightarrow{p} 0$ and $b_n = \frac{1}{n}\Sigma_{i=1}^{n} X_i^r u_i^s$ where r and s are integers. For example, $a_n = (\overline{X} - \mu_X)$, $a_n = (\hat{\beta}_1 - \beta_1)$ and so forth. The result then follows from Slutksy's theorem if $\frac{1}{n}\Sigma_{i=1}^{n} X_i^r u_i^s \xrightarrow{p} d$ where d is a finite constant. Let $w_i = X_i^r u_i^s$ and note that w_i is i.i.d. The law of large numbers can then be used for the desired result if $E(w_i^2) < \infty$. There are two cases that need to be addressed. In the first, both r and s are non-zero. In this case write

$$E(w_i^2) = E(X_i^{2r} u_i^{2s}) < \sqrt{[E(X_i^{4r})][E(u_i^{4s})]}$$

and this term is finite if r and s are less than 2. Inspection of the terms shows that this is true. In the second case, either $r = 0$ or $s = 0$. In this case the result follows directly if the non-zero exponent (r or s) is less than 4. Inspection of the terms shows that this is true.

Chapter 18
The Theory of Multiple Regression

Solutions to Exercises

1. (a) The regression in the matrix form is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}$$

with

$$\mathbf{Y} = \begin{pmatrix} TestScore_1 \\ TestScore_2 \\ \vdots \\ TestScore_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & Income_1 & Income_1^2 \\ 1 & Income_2 & Income_2^2 \\ \vdots & \vdots & \vdots \\ 1 & Income_n & Income_n^2 \end{pmatrix}$$

$$\mathbf{U} = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}.$$

(b) The null hypothesis is

$$\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$$

versus $\mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}$ with

$$\mathbf{R} = (0 \quad 0 \quad 1) \text{ and } \mathbf{r} = 0.$$

The heteroskedasticity-robust F-statistic testing the null hypothesis is

$$F = (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})'\left[\mathbf{R}\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}}\mathbf{R}'\right]^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})/q$$

With $q = 1$. Under the null hypothesis,

$$F \xrightarrow{d} Fq, \infty.$$

We reject the null hypothesis if the calculated F-statistic is larger than the critical value of the $F_{q,\infty}$ distribution at a given significance level.

3. (a)

$$\begin{aligned}\text{Var}(Q) &= E[(Q-\mu_Q)^2] \\ &= E[(Q-\mu_Q)(Q-\mu_Q)'] \\ &= E[(\mathbf{c}'\mathbf{W}-\mathbf{c}'\boldsymbol{\mu}_\mathbf{W})(\mathbf{c}'\mathbf{W}-\mathbf{c}'\boldsymbol{\mu}_\mathbf{W})'] \\ &= \mathbf{c}'E[(\text{W}-\boldsymbol{\mu}_\mathbf{W})(\mathbf{W}-\boldsymbol{\mu}_\mathbf{W})']\mathbf{c} \\ &= \text{c}'\,\text{var}(\mathbf{W})\,\text{c} = \mathbf{c}'\boldsymbol{\Sigma}_\mathbf{w}\mathbf{c}\end{aligned}$$

where the second equality uses the fact that Q is a scalar and the third equality uses the fact that $\mu_Q = \mathbf{c}'\boldsymbol{\mu}_\mathbf{w}$.

(b) Because the covariance matrix $\sum_\text{W}$ is positive definite, we have $\mathbf{c}'\sum_\mathbf{w}\mathbf{c} > 0$ for every non-zero vector from the definition. Thus, var(Q) > 0. Both the vector **c** and the matrix $\sum_\text{W}$ are finite, so var(Q) = $\mathbf{c}'\sum_\mathbf{w}\mathbf{c}$ is also finite. Thus, 0 < var(Q) < ∞.

5. $\mathbf{P_X} = \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'}$, $\mathbf{M_X} = \mathbf{I}_n - \mathbf{P_X}$.

(a) $\mathbf{P_X}$ is idempotent because

$$\mathbf{P_XP_X} = \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'X}(\mathbf{X'X})^{-1}\mathbf{X'} = \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'} = \mathbf{P_X}.$$

$\mathbf{M_X}$ is idempotent because

$$\begin{aligned}\mathbf{M_XM_X} &= (\mathbf{I}_n-\mathbf{P_X})(\mathbf{I}_n-\mathbf{P_X}) = \mathbf{I}_n-\mathbf{P_X}-\mathbf{P_X}+\mathbf{P_XP_X} \\ &= \mathbf{I}_n - 2\mathbf{P_X}+\mathbf{P_X} = \mathbf{I_n}-\mathbf{P_X} = \mathbf{M_X}\end{aligned}$$

$\mathbf{P_XM_X} = 0_{nxn}$ because

$$\mathbf{P_XM_X} = \mathbf{P_X}(\mathbf{I}_\text{n} - \mathbf{P_X}) = \mathbf{P_X} - \mathbf{P_XP_X} = \mathbf{P_X} - \mathbf{P_X} = 0_{n\times n}$$

(b) Because $\hat{\beta} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$, we have

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'Y} = \mathbf{P_XY}$$

which is Equation (18.27). The residual vector is

$$\hat{\mathbf{U}} = \mathbf{Y}-\hat{\mathbf{Y}} = \mathbf{Y}-\mathbf{P_XY} = (\mathbf{I}_n-\mathbf{P_X})\mathbf{Y} = \mathbf{M_XY}.$$

We know that $\mathbf{M_XX}$ is orthogonal to the columns of **X**:

$$\mathbf{M_XX} = (\mathbf{I}_n - \mathbf{P_X})\mathbf{X} = \mathbf{X} - \mathbf{P_XX} = \mathbf{X} - \mathbf{X}(\mathbf{X'X})^{-1}\mathbf{X'X} = \mathbf{X} - \mathbf{X} = \mathbf{0}$$

so the residual vector can be further written as

$$\hat{\mathbf{U}} = \mathbf{M_XY} = \mathbf{M_X}(\mathbf{X}\boldsymbol{\beta}+\mathbf{U}) = \mathbf{M_XX}\boldsymbol{\beta}+\mathbf{M_XU} = \mathbf{M_XU}$$

which is Equation (18.28).

7. (a) We write the regression model, $Y_i = \beta_1X_i + \beta_2W_i + u_i$, in the matrix form as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\gamma} + \mathbf{U}$$

with

$$\mathbf{Y}=\begin{pmatrix}Y_1\\Y_2\\\vdots\\Y_n\end{pmatrix},\qquad \mathbf{X}=\begin{pmatrix}X_1\\X_2\\\vdots\\X_n\end{pmatrix},\qquad \mathbf{W}=\begin{pmatrix}W_1\\W_2\\\vdots\\W_n\end{pmatrix},\qquad \mathbf{U}=\begin{pmatrix}u_1\\u_2\\\vdots\\u_n\end{pmatrix},$$

$$\boldsymbol{\beta}=\beta_1,\qquad \gamma=\beta_2.$$

The OLS estimator is

$$\begin{aligned}\begin{pmatrix}\hat\beta_1\\\hat\beta_2\end{pmatrix}&=\begin{pmatrix}\mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{W}\\\mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{W}\end{pmatrix}^{-1}\begin{pmatrix}\mathbf{X}'\mathbf{Y}\\\mathbf{W}'\mathbf{Y}\end{pmatrix}\\
&=\begin{pmatrix}\beta_1\\\beta_2\end{pmatrix}+\begin{pmatrix}\mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{W}\\\mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{W}\end{pmatrix}^{-1}\begin{pmatrix}\mathbf{X}'\mathbf{U}\\\mathbf{W}'\mathbf{U}\end{pmatrix}\\
&=\begin{pmatrix}\beta_1\\\beta_2\end{pmatrix}+\begin{pmatrix}\frac{1}{n}\mathbf{X}'\mathbf{X} & \frac{1}{n}\mathbf{X}'\mathbf{W}\\\frac{1}{n}\mathbf{W}'\mathbf{X} & \frac{1}{n}\mathbf{W}'\mathbf{W}\end{pmatrix}^{-1}\begin{pmatrix}\frac{1}{n}\mathbf{X}'\mathbf{U}\\\frac{1}{n}\mathbf{W}'\mathbf{U}\end{pmatrix}\\
&=\begin{pmatrix}\beta_1\\\beta_2\end{pmatrix}+\begin{pmatrix}\frac{1}{n}\sum_{i=1}^n X_i^2 & \frac{1}{n}\sum_{i=1}^n X_iW_i\\\frac{1}{n}\sum_{i=1}^n W_iX_i & \frac{1}{n}\sum_{i=1}^n W_i^2\end{pmatrix}^{-1}\begin{pmatrix}\frac{1}{n}\sum_{i=1}^n X_iu_i\\\frac{1}{n}\sum_{i=1}^n W_iu_i\end{pmatrix}\end{aligned}$$

By the law of large numbers $\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{p} E(X^2)$; $\frac{1}{n}\sum_{i=1}^n W_i^2 \xrightarrow{p} E(W^2)$; $\frac{1}{n}\sum_{i=1}^n X_iW_i \xrightarrow{p} E(XW)=0$ (because X and W are independent with means of zero); $\frac{1}{n}\sum_{i=1}^n X_iu_i \xrightarrow{p} E(Xu)=0$ (because X and u are independent with means of zero); $\frac{1}{n}\sum_{i=1}^n X_iu_i \xrightarrow{p} E(Xu)=0$ Thus

$$\begin{aligned}\begin{pmatrix}\hat\beta_1\\\hat\beta_2\end{pmatrix}&\xrightarrow{p}\begin{pmatrix}\beta_1\\\beta_2\end{pmatrix}+\begin{pmatrix}E(X^2) & 0\\0 & E(W^2)\end{pmatrix}^{-1}\begin{pmatrix}0\\E(Wu)\end{pmatrix}\\
&=\begin{pmatrix}\beta_1\\\beta_2+\frac{E(Wu)}{E(W^2)}\end{pmatrix}.\end{aligned}$$

(b) From the answer to (a) $\hat\beta_2 \xrightarrow{p} \beta_2+\frac{E(Wu)}{E(W^2)}\neq\beta_2$ if $E(Wu)$ is nonzero.

(c) Consider the population linear regression u_i onto W_i:

$$u_i = \lambda W_i + a_i$$

where $\lambda = E(Wu)/E(W^2)$. In this population regression, by construction, $E(aW) = 0$. Using this equation for u_i rewrite the equation to be estimated as

$$\begin{aligned}Y_i&=X_i\beta_1+W_i\beta_2+u_i\\&=X_i\beta_1+W_i(\beta_2+\lambda)+a_i\\&=X_i\beta_1+W_i\theta+a_i\end{aligned}$$

where $\theta=\beta_2+\lambda.$ A calculation like that used in part (a) can be used to show that

$$\begin{pmatrix} \sqrt{n}(\hat{\beta}_1 - \beta_1) \\ \sqrt{n}(\hat{\beta}_2 - \theta) \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_i^2 & \frac{1}{n}\sum_{i=1}^n X_i W_i \\ \frac{1}{n}\sum_{i=1}^n W_i X_i & \frac{1}{n}\sum_{i=1}^n W_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i a_i \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n W_u a_i \end{pmatrix}$$

$$\xrightarrow{d} \begin{pmatrix} E(X^2) & 0 \\ 0 & E(W^2) \end{pmatrix}^{-1} \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}$$

where S_1 is distributed $N(0, \sigma_a^2 E(X_2))$. Thus by Slutsky's theorem

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) \xrightarrow{d} N\left(0,\ \frac{\sigma_a^2}{E(X^2)}\right)$$

Now consider the regression that omits W, which can be written as:

$$Y_i = X_i\beta_1 + d_i$$

where $d_i = W_i\theta + a_i$. Calculations like those used above imply that

$$\sqrt{n}\left(\hat{\beta}_1^r - \beta_1\right) \xrightarrow{d} N\left(0,\ \frac{\sigma_d^2}{E(X^2)}\right).$$

Since $\sigma_d^2 = \sigma_a^2 + \theta^2 E(W^2)$, the asymptotic variance of $\hat{\beta}_1^r$ is never smaller than the asymptotic variance of $\hat{\beta}_1$.

9. (a)

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{M_W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{M_W}\mathbf{Y} \\ &= (\mathbf{X}'\mathbf{M_W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{M_W}(\mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\gamma} + \mathbf{U}) \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{M_W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{M_W}\mathbf{U}. \end{aligned}$$

The last equality has used the orthogonality $\mathbf{M_W W = 0}$. Thus

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{M_W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{M_W}\mathbf{U} = (n^{-1}\mathbf{X}'\mathbf{M_W}\mathbf{X})^{-1}(n^{-1}\mathbf{X}'\mathbf{M_W}\mathbf{U}).$$

(b) Using $\mathbf{M_W} = \mathbf{I}_n - \mathbf{P_W}$ and $\mathbf{P_W} = \mathbf{W}(\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'$ we can get

$$\begin{aligned} n^{-1}\mathbf{X}'\mathbf{M_W}\mathbf{X} &= n^{-1}\mathbf{X}'(\mathbf{I}_n - \mathbf{P_W})\mathbf{X} \\ &= n^{-1}\mathbf{X}'\mathbf{X} - n^{-1}\mathbf{X}'\mathbf{P_W}\mathbf{X} \\ &= n^{-1}\mathbf{X}'\mathbf{X} - (n^{-1}\mathbf{X}'\mathbf{W})(n^{-1}\mathbf{W}'\mathbf{W})^{-1}(n^{-1}\mathbf{W}'\mathbf{X}). \end{aligned}$$

First consider $n^{-1}\mathbf{X}'\mathbf{X} = \frac{1}{n}\sum_{i=1}^n \mathbf{X}_i\mathbf{X}_i'$. The (j, l) element of this matrix is $\frac{1}{n}\sum_{i=1}^n \mathbf{X}_{ji}\mathbf{X}_{li}$. By Assumption (ii), $\mathbf{X}_i$ is i.i.d., so $X_{ji}X_{li}$ is i.i.d. By Assumption (iii) each element of $\mathbf{X}_i$ has four moments, so by the Cauchy-Schwarz inequality $X_{ji}X_{li}$ has two moments:

$$E(X_{ji}^2 X_{li}^2) \le \sqrt{E(X_{ji}^4) \cdot E(X_{li}^4)} < \infty.$$

Because $X_{ji}X_{li}$ is i.i.d. with two moments, $\frac{1}{n}\sum_{i=1}^n X_{ji}X_{li}$ obeys the law of large numbers, so

$$\frac{1}{n}\sum_{i=1}^{n} X_{ji}X_{li} \xrightarrow{p} E(X_{ji}X_{li}).$$

This is true for all the elements of $n^{-1}\mathbf{X}'\mathbf{X}$, so

$$n^{-1}\mathbf{X}'\mathbf{X} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i' \xrightarrow{p} E(\mathbf{X}_i\mathbf{X}_i') = \Sigma_{\mathbf{XX}}.$$

Applying the same reasoning and using Assumption (ii) that ($\mathbf{X}_i$, $\mathbf{W}_i$, Y_i) are i.i.d. and Assumption (iii) that ($\mathbf{X}_i$, $\mathbf{W}_i$, u_i) have four moments, we have

$$n^{-1}\mathbf{W}'\mathbf{W} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{W}_i\mathbf{W}_i' \xrightarrow{p} E(\mathbf{W}_i\mathbf{W}_i') = \Sigma_{\mathbf{WW}},$$

$$n^{-1}\mathbf{X}'\mathbf{W} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{W}_i' \xrightarrow{p} E(\mathbf{X}_i\mathbf{W}_i') = \Sigma_{\mathbf{XW}},$$

and

$$n^{-1}\mathbf{W}'\mathbf{X} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{W}_i\mathbf{X}_i' \xrightarrow{p} E(\mathbf{W}_i\mathbf{X}_i') = \Sigma_{\mathbf{WX}}.$$

From Assumption (iii) we know $\Sigma_{\mathbf{XX}}, \Sigma_{\mathbf{WW}}, \Sigma_{\mathbf{XW}}$, and $\Sigma_{\mathbf{WX}}$ are all finite non-zero, Slutsky's theorem implies

$$\begin{aligned} n^{-1}\mathbf{X}'\mathbf{M}_{\mathbf{W}}\mathbf{X} &= n^{-1}\mathbf{X}'\mathbf{X} - (n^{-1}\mathbf{X}'\mathbf{W})(n^{-1}\mathbf{W}'\mathbf{W})^{-1}(n^{-1}\mathbf{W}'\mathbf{X}) \\ &\xrightarrow{p} \mathbf{\Sigma}_{\mathbf{XX}} - \mathbf{\Sigma}_{\mathbf{XW}}\mathbf{\Sigma}_{\mathbf{WW}}^{\mathbf{-1}}\mathbf{\Sigma}_{\mathbf{WX}} \end{aligned}$$

which is finite and invertible.

(c) The conditional expectation

$$E(\mathbf{U}|\mathbf{X},\mathbf{W}) = \begin{pmatrix} E(u_1|\mathbf{X},\mathbf{W}) \\ E(u_2|\mathbf{X},\mathbf{W}) \\ \vdots \\ E(u_n|\mathbf{X},\mathbf{W}) \end{pmatrix} = \begin{pmatrix} E(u_1|\mathbf{X}_1,\mathbf{W}_1) \\ E(u_2|\mathbf{X}_2,\mathbf{W}_2) \\ \vdots \\ E(u_n|\mathbf{X}_n,\mathbf{W}_n) \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{W}_1'\boldsymbol{\delta} \\ \mathbf{W}_2'\boldsymbol{\delta} \\ \vdots \\ \mathbf{W}_n'\boldsymbol{\delta} \end{pmatrix} = \begin{pmatrix} \mathbf{W}_1' \\ \mathbf{W}_2' \\ \vdots \\ \mathbf{W}_n' \end{pmatrix}\boldsymbol{\delta} = \mathbf{W}\boldsymbol{\delta}.$$

The second equality used Assumption (ii) that $(\mathbf{X}_i, \mathbf{W}_i, Y_i)$ are i.i.d., and the third equality applied the conditional mean independence assumption (i).

(d) In the limit

$$n^{-1}\mathbf{X}'\mathbf{M}_{\mathbf{W}}\mathbf{U} \xrightarrow{p} E(\mathbf{X}'\mathbf{M}_{\mathbf{W}}\mathbf{U}|\mathbf{X},\mathbf{W}) = \mathbf{X}'\mathbf{M}_{\mathbf{W}}E(\mathbf{U}|\mathbf{X},\mathbf{W}) = \mathbf{X}'\mathbf{M}_{\mathbf{W}}\mathbf{W}\boldsymbol{\delta} = \mathbf{0}_{k_1\times 1}$$

because $\mathbf{M_W W} = \mathbf{0}$.

(e) $n^{-1}\mathbf{X}'\mathbf{M_W X}$ converges in probability to a finite invertible matrix, and $n^{-1}\mathbf{X}'\mathbf{M_W U}$ converges in probability to a zero vector. Applying Slutsky's theorem,

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (n^{-1}\mathbf{X}'\mathbf{M_W X})^{-1}(n^{-1}\mathbf{X}'\mathbf{M_W U}) \xrightarrow{p} \mathbf{0}.$$

This implies

$$\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}.$$

11 (a) Using the hint $\boldsymbol{C} = [\boldsymbol{Q}_1\,\boldsymbol{Q}_2]\begin{bmatrix} \boldsymbol{I}_r & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \boldsymbol{Q}_1' \\ \boldsymbol{Q}_2' \end{bmatrix}$, where $\boldsymbol{Q}'\boldsymbol{Q} = \boldsymbol{I}$. The result follows with $\boldsymbol{A}=\boldsymbol{Q}_1$.

(b) $\boldsymbol{W} = \boldsymbol{A}'\boldsymbol{V} \sim \mathrm{N}(\boldsymbol{A}'\boldsymbol{0}, \boldsymbol{A}'\boldsymbol{I}_n\boldsymbol{A})$ and the result follows immediately.

(c) $\boldsymbol{V}'\boldsymbol{C}\boldsymbol{V} = \boldsymbol{V}'\boldsymbol{A}\boldsymbol{A}'\boldsymbol{V} = (\boldsymbol{A}'\boldsymbol{V})'(\boldsymbol{A}'\boldsymbol{V}) = \boldsymbol{W}'\boldsymbol{W}$ and the result follows from (b).

13 (a) This follows from the definition of the Lagrangian.

(b) The first order conditions are

$$(*)\ \boldsymbol{X}'(\boldsymbol{Y}-\boldsymbol{X}\tilde{\boldsymbol{\beta}}) + \boldsymbol{R}'\boldsymbol{\lambda} = 0$$

and

$$(**)\ \boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r} = 0$$

Solving (*) yields

$$(***)\ \tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\boldsymbol{\lambda}.$$

Multiplying by $\boldsymbol{R}$ and using (**) yields $\boldsymbol{r} = \boldsymbol{R}\hat{\boldsymbol{\beta}} + \boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\boldsymbol{\lambda}$, so that

$$\lambda = -[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}).$$

Substituting this into (***) yields the result.

(c) Using the result in (b), $\boldsymbol{Y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}} = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) + \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})$, so that

$$(\boldsymbol{Y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}) = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) + (\boldsymbol{R}\hat{\boldsymbol{\beta}} - r)'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})$$
$$+ 2(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}).$$

But $(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'\boldsymbol{X} = 0$, so the last term vanishes, and the result follows.

(d) The result in (c) shows that $(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}) = SSR_{Restricted} - SSR_{Unrestricted}$. Also $s_u^2 = SSR_{Unrestricted}/(n - k_{Unrestricted} - 1)$, and the result follows immediately.

15 (a) This follows from exercise (18.6).

(b) $\tilde{\boldsymbol{Y}}_i = \tilde{\boldsymbol{X}}_i\beta + \tilde{\boldsymbol{u}}_i$, so that

$$\hat{\beta}-\beta=\left(\sum_{i=1}^{n}\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i\right)^{-1}\sum_{i=1}^{n}\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{u}}_i$$
$$=\left(\sum_{i=1}^{n}\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i\right)^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i'\boldsymbol{M}'\boldsymbol{M}\boldsymbol{u}_i$$
$$=\left(\sum_{i=1}^{n}\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i\right)^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i'\boldsymbol{M}'\boldsymbol{u}_i$$
$$=\left(\sum_{i=1}^{n}\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i\right)^{-1}\sum_{i=1}^{n}\tilde{\boldsymbol{X}}_i'\boldsymbol{u}_i$$

(c) Note Typo in problem: Should Read: $\boldsymbol{Q}_{\tilde{X}} = T^{-1}E(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i) = T^{-1}\sum_{t=1}^{T}E(X_{it}-\overline{X}_i)^2$
$\hat{\boldsymbol{Q}}_{\tilde{X}} = \frac{1}{n}\sum_{i=1}^{n}(T^{-1}\sum_{t=1}^{T}(X_{it}-\overline{X}_i)^2)$, where $(T^{-1}\sum_{t=1}^{T}(X_{it}-\overline{X}_i)^2)$ are i.i.d. with mean $\boldsymbol{Q}_{\tilde{X}}$ and finite variance (because X_{it} has finite fourth moments). The result then follows from the law of large numbers.

(d) This follows the the Central limit theorem.

(e) This follows from Slutsky's theorem.

(f) η_i^2 are i.i.d., and the result follows from the law of large numbers.

(g) Let $\hat{\eta}_i = T^{-1/2}\tilde{\boldsymbol{X}}_i'\hat{\tilde{\boldsymbol{u}}}_i = \eta_i - T^{-1/2}(\hat{\beta}-\beta)\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i$. Then
$\hat{\eta}_i^2 = T^{-1/2}\tilde{\boldsymbol{X}}_i'\hat{\tilde{\boldsymbol{u}}}_i = \eta_i^2 + T^{-1}(\hat{\beta}-\beta)^2(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i)^2 - 2T^{-1/2}(\hat{\beta}-\beta)\eta_i\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i$
and $\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_i^2 - \frac{1}{n}\sum_{i=1}^{n}\eta_i^2 = T^{-1}(\hat{\beta}-\beta)^2\frac{1}{n}\sum_{i=1}^{n}(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i)^2 - 2T^{-1/2}(\hat{\beta}-\beta)\frac{1}{n}\sum_{i=1}^{n}\eta_i\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i$

Because $(\hat{\beta}-\beta)\xrightarrow{p}0$, the result follows from (a) $\frac{1}{n}\sum_{i=1}^{n}(\tilde{X}_i'\tilde{X}_i)^2\xrightarrow{p}E[(\tilde{X}_i'\tilde{X}_i)^2]$ and (b) $\frac{1}{n}\sum_{i=1}^{n}\eta_i\tilde{X}_i'\tilde{X}_i\xrightarrow{p}E(\eta_i\tilde{X}_i'\tilde{X}_i)$. Both (a) and (b) follow from the law of large numbers; both (a) and (b) are averages of i.i.d. random variables. Completing the proof requires verifying that $(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i)^2$ has two finite moments and $\eta_i\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i$ has two finite moments. These in turn follow from 8-moment assumptions for (X_{it}, u_{it}) and the Cauchy-Schwartz inequality. Alternatively, a "strong" law of large numbers can be used to show the result with finite fourth moments.